Docket No.: MXI-211
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Debra Hudson *et al.*

Application No.: 10/073644 Confirmation No.: 6293

Filed: February 11, 2002 Art Unit: 1644

For: HUMAN MONOCLONAL ANTIBODIES TO FC ALPHA RECEPTOR (CD89) Examiner: M. A. Belyavskyi

DECLARATION BY DR. NILS LONBERG UNDER 37 C.F.R. §1.132

MS AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

I, Dr. Nils Lonberg, declare the following:

1. I, Dr. Nils Lonberg, am presently the Senior Vice President and Scientific Director at Medarex, Inc. in Milpitas, California, the assignee of the above-referenced patent application. I received a Ph.D. in Biochemistry and Molecular Biology from Harvard University and completed a Post-Doctoral Fellowship at Memorial Sloan-Kettering Cancer Center in New York, New York. My *curriculum vitae* is attached herewith as Appendix A.

2. I have reviewed claims 64-66 of the above-referenced application which are drawn to an isolated human monoclonal anti-human CD89 antibody, or antigen binding portion thereof, comprising a heavy chain variable region derived from a human germline V_H 3-30.3 gene and a light chain variable region derived from either a human germline V_K L18 gene or V_K A27 gene.

3. I understand that claims 64-66 of the above-referenced application have been rejected as being indefinite. Specifically, the Examiner asserts that claims 64-66 are indefinite based on reference to the V_H 3-30.3, V_K L18, and V_K A27 germline genes "because the characteristics of these genes are not known."

4. It is my opinion that, prior to the filing date of the present application, the meaning of the above-mentioned human germline genes would have been clear and definite to one of ordinary skill in the art, including the specific characteristics (*e.g.*, sequences) of these genes. In particular, as evidenced by the enclosed references (discussed in detail below), not only was the

nomenclature of these genes well established, accepted and known in the art by the filing date of the application, but also they had been mapped and sequenced in their entirety. In addition, as also evidenced by the enclosed references, each of these V_H 3-30.3, V_K L18, and V_K A27 designations corresponds to a single gene (*i.e.*, a single allele), the full-length sequence of which was known and publicly available at the filing date of the present application. As such, use of these gene designations in the claims of the present application would have been clear and definite to one of ordinary skill in the art.

5. With respect to the human light chain germline genes, V_K L18 and V_K A27, I refer to the following scientific review entitled "Immunoglobulin Genes," Second Ed., (1995) edited by T. Honjo and F.W. Alt, Academic Press, which summarizes the knowledge in the art as of 1995 with respect to the organization, structure and nomenclature of the light chain immunoglobulin genes. In particular, chapter 8, entitled "The human immunoglobulin κ genes" by Hans Zachau, discusses the human kappa locus (see pages 173-191; attached herewith as Appendix B). As described in Appendix B, the human V_K genes were first isolated and sequenced as early as 1980. An outline of the human kappa locus is provided in Figure 2 (page 175) and shows that the κ proteins have been classified into four subgroups, I-IV with the V_K L18 gene and V_K A27 genes classified in subgroups I and III, respectively. Moreover, as also shown in Figure 2, these genes correspond to a single allele. Moreover, the complete sequences for these genes had been published, for example, by Schable and Zachau (1993) *Biol. Chem.* 374:1001-1022 (attached herewith as Appendix C; see page 1020 providing the amino acid sequences of V_K L18 and V_K A27).

6. With respect to the human heavy chain germline gene V_H 3-30.3, I refer again to the Honjo review. In particular, chapter 7, entitled "Immunoglobulin heavy chain loci of mouse and human" by Tasuku Honjo and Fumihiko Matsuda, discusses the nomenclature of the human V_H locus (see pages 145-171; attached herewith as Appendix D). The human V_H regions have been divided into three subgroups based on amino acid sequence homology. These three subgroups have further been divided into six V_H families (see Figure 1, page 147). Subgroup III contains the largest number of members, yet constitutes a single family, V_H3. Based on the established nomenclature, each V_H segment is named according to its family number and the order from the 3' end of the V_H locus. An insertional polymorphic V_H segment is indicated by a number with a decimal point (see, page 149, lines 22-27). Accordingly, the V_H 3-30.3 gene refers to a specific polymorphic variant of the V_H 3 family member which is the thirtieth gene from the D region. In addition, the complete amino acid sequence of allele 3 of V_H 3-30, *i.e.*, V_H 3-30.3, had been determined and published prior to the filing date of the present application by, for example, Chang and Siegel (1998) *Am. Soc. Hematol.* 21(8)3066-3078 (attached herewith as Appendix E; see Figure 2(a)).

7. In conclusion, as shown by the foregoing pre-filing publications, the light and heavy chain immunoglobulin germline gene nomenclature recited in the claims of the present application was well known and accepted in the art prior to the filing date of the present application. Moreover, these gene designations (V_H 3-30.3, V_K L18, and V_K A27) were known to correspond to a single gene sequence (*i.e.*, a single allele) which was also publicly available prior to the filing date of the application. Accordingly, reference to these gene designations in the present claims would have been clear and definite to one of ordinary skill in the art at the time of filing.

9. I have been warned that willful false statements and the like so made are punishable by fine or imprisonment, or both, under §1001 of Title 18 or the United States Code, and that such willful and false statements may jeopardize the validity of the subject application or any patent resulting therefrom, and declare that all statements made of our own knowledge are true and that all statements made on information and belief are believed to be true.

By: _____
Nils Lonberg, Ph.D.

Date: 10-13-05

Appendix A

Curriculum Vitae

Name: **Nils Lonberg**

Born: 2-23-56, Berkeley, CA

Social Security #: 227-92-3993

Home Address: 168 Bardet Rd
Woodside, CA 94062

Work Address: Medarex
521 Cottonwood Drive
Milpitas, CA 95035

Telephone: home: (650) 851-1442
work: (408) 545-2721

Internet E-mail: nlonberg@medarex.com

Education:

Undergraduate:	1974-1977	Reed College
	1977-1978	University of Oregon
	1978-1979	Reed College
Graduate:	1979-1985	Harvard University, Ph.D., March 1985
Postdoctoral:	1985-1989	Sloan-Kettering Institute for Cancer Research

Employment:

1990-1994	Senior Scientist GenPharm International, Mountain View, CA.
1994-1997	Director, Molecular Biology GenPharm International, Palo Alto, CA.
1998-2000	Vice President, Scientific Director GenPharm International, San Jose, CA.
2000-present	Senior Vice President, Scientific Director Medarex, Milpitas, CA.

Selected Bibliography:

1. **S.C. Kowalczykowski, N. LONBERG, J.W. Newport, and P.H. von Hippel** (1981) Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids, I. Characterization of the binding interactions. *J. Mol. Biol.* **145**, 75-104.

2. **J.W. Newport, N. LONBERG, S.C. Kowalczykowski, and P.H. von Hippel** (1981) Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids, II. Specificity of binding to DNA and RNA. *J. Mol. Biol.* **145**, 105-121.

3. **N. LONBERG, S.C. Kowalczykowski, L.S. Paul, and P.H. von Hippel** (1981) Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids, III. Binding properties of two specific proteolytic digestion products of the protein (G32P*I and G32P*III). *J. Mol. Biol.* **145**, 123-138.

4. **P.H. von Hippel, S.C. Kowalczykowski, N. LONBERG, J.W. Newport, L.S. Paul, G.D. Stormo, and L. Gold.** (1982) Autoregulation of gene expression, quantitative evaluation of the expression and function of the bacteriophage T4 gene 32 (single-stranded DNA binding protein) protein system. *J. Mol. Biol.* **162**, 795-818.

5. **N. LONBERG and W. Gilbert** (1983) Primary structure of chicken muscle pyruvate kinase mRNA. *Proc. Natl. Acad. Sci. USA* **80**, 3661-3665.

6. **N. LONBERG and W. Gilbert** (1985) Intron/exon structure of the chicken pyruvate kinase gene. *Cell* **40**, 81-90.

7. **S.C. Kowalczykowski, L.S. Paul, N. LONBERG, J.W. Newport, J.A. McSwiggen and P.H. von Hippel** (1986) The cooperative and noncooperative binding of protein ligands to nucleic acid lattices: Experimental approaches to the determination of thermodynamic parameters. *Biochemistry* **25**, 1226-1240.

8. **N. LONBERG, S.N. Gettner, E. Lacy, and D.R. Littman** (1988) Mouse brain CD4 transcripts encode only the COOH-terminal half of the protein. *Mol. Cell. Biol.* **8**, 2224-2228.

9. **A.M. Norment, N. LONBERG, E. Lacy, and D.R. Littman** (1989) Alternatively spliced mRNA encodes a secreted form of human CD8α: Characterization of the human CD8α gene. *J. Immunol.* **142**, 3312-3319.

10. **H. Clevers, N. LONBERG, S. Dunlap, E. Lacy, and C. Terhorst** (1989) An enhancer located in a CpG-island 3' of the TCR/CD3-ε gene confers T lymphocyte-specificity to its promoter. *EMBO J.* **8**, 2527-2535.

11. **H. Meade, L. Gates, E. Lacy, and N. LONBERG** (1990) Bovine α-S_1-casein gene sequences direct high level expression of active human urokinase in mouse milk. *Bio/Technology.* **8**, 443-446.

12. **R.S. Blumberg, S. Ley, J. Sancho, N. LONBERG, E. Lacy, F. McDermott, V. Schad, J.L. Greenstein, and C. Terhorst** (1990) Structure of the T cell antigen receptor: Evidence for two CD3-ε subunits in the T cell receptor/CD3 complex. *Proc. Natl. Acad. Sci. USA.* **87**, 7220-7224.

13. **L.D. Taylor, C.E. Carmack, S.R. Schramm, R. Mashayekh, K.M. Higgins, C.-C. Kuo, C. Woodhouse, R.M. Kay, and N. LONBERG** (1992) A transgenic mouse that expresses a diversity of human sequence heavy and light chain immunoglobulins. *Nucleic Acids Res.* **20**, 6287-6295.

14. **N. Tuaillon, L.D. Taylor, N. LONBERG, P.W. Tucker, and J.D. Capra** (1993) Human immunoglobulin heavy-chain minilocus recombination in transgenic mice: Gene-segment use in μ and γ transcripts. *Proc. Natl. Acad. Sci. USA.* **90**, 3720-3724.

15. **L.D. Taylor, C.E. Carmack, D. Huszar, K.M. Higgins, R. Mashayekh, G. Sequar, S.R. Schramm, C.-C. Kuo, S.L. O'Donnell, R.M. Kay, C.S. Woodhouse, and N. LONBERG** (1994) Human immunoglobulin transgenes undergo rearrangement, somatic mutation, and class switching in mice that lack endogenous IgM. *Int. Immunol.* **6**, 579-591.

16. **N. LONBERG, L.D. Taylor, F.A. Harding, M. Trounstine, K.M. Higgins, S.R. Schramm, C.-C. Kuo, R. Mashayekh, K. Wymore, J.G. McCabe, D. Munoz-O'Regan, S. L. O'Donnell, E.S.G. Lapachet, T. Bengoechea, D.M. Fishwild, C.E. Carmack, R.M. Kay, and D. Huszar** (1994) Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature,* **368**, 856-859.

17. **N. LONBERG** (1994) Transgenic approaches to human monclonal antibodies. in *Handbook of Experimental Pharmacology,* Vol. 113, The Pharmacology of Monoclonal Antibodies, Eds. M. Rosenberg, G. Moore (Springer-Verlag, New York), pp. 49-101.

18. **N. Tuaillon, A.B. Miller, N. LONBERG, P.W. Tucker, and J.D. Capra** (1994) Biased utilization of DHQ52 and JH4 gene segments in a human Ig transgenic minilocus is independent of antigenic selection. *J. Immunol.* **152**, 2912-2920.

19. **B. Wang, C. Biron, J. She, K. Higgins, M.-J. Sunshine, E. Lacy, N. LONBERG, and C. Terhorst** (1994) A block in early T lymphocyte and natural killer cell development in transgenic mice with high-copy numbers of the human CD3ε gene. *Proc. Natl. Acad. Sci. USA* **91**, 9402-9406.

20. **N. LONBERG and D. Huszar** (1995) Human antibodies from transgenic mice. *Int. Rev. Immunol.* **13**, 65-93.

21. **F.A. Harding and N. LONBERG** (1995) Class switching in human immunoglobulin transgenic mice. *Ann. New York Acad Sci.* **764**, 536-546.

22. **B. Wang, C. Levelt, M. Salio, D. Zheng, J. Sancho, C.-P. Liu, J. She, M. Huang, K. Higgins, M.-J. Sunshine, K. Eichmann, E. Lacy, N. LONBERG, and C. Terhorst** (1995) Over-expression of CD3ε transgenes blocks T lymphocyte development. *Int. Immunol.* **7**, 435-448.

23. **D.M. Fishwild, S.L. O'Donnell, T. Bengoechea, D.V. Hudson, F. Harding, S.L. Bernhard, D. Jones, R.M. Kay, K.M. Higgins, S.R. Schramm and N. LONBERG** (1996) High avidity human IgGκ monoclonal antibodies from a novel strain of minilocus transgenic mice. *Nature Biotechnology* **14**, 845-851.

24. **X. Wang, E.A. Bornslaeger, O. Haub, C. Tomihara-Newberger, N. LONBERG, M.B. Dinulos, C.M. Disteche, N. Copeland, D.J. Gilbert, N.A. Jenkins, and E. Lacy** (1996) A candidate gene for the *amnionless* gastrulation stage mouse mutation encodes a TRAF related protein. *Dev. Biol.* **177**, 274-290.

25. **B. Wang, J. She, M. Salio, D. Allen, E. Lacy, N. LONBERG, C. Terhorst** (1997) CD3-ε overexpressed in prothymocytes acts as an oncogene. *Molecular Medicine,* **3**, 72-81.

26. **M.E. Donohoe, G.B. Beck-Engeser, N. LONBERG, H. Karasuyama, R.L. Riley, H.M. Jack, B.B. Blomberg** (2000) Transgenic human lambda 5 rescues the murine lambda 5 nullizygous phenotype. *J. Immunol.* **10**, 5269-76.

27. **I. Ishida, K. Tomizuka, H. Yoshida, T. Tahara, N. Takahashi, A. Ohguma, S. Tanaka, M. Umehashi, H. Maeda, C. Nozaki, E. Halk, N. LONBERG** (2002) Production of human monoclonal and polyclonal antibodies in TransChromo animals. *Cloning Stem Cells.* **4**, 91-102.

28. **T. Keler, E. Halk, L. Vitale, T. O'Neill, D. Blanset, S. Lee, M. Srinivasan, R.F. Graziano, T. Davis, N. LONBERG, A. Korman** (2003) Activity and safety of CTLA-4 blockade combined with vaccines in cynomolgus macaques. *J Immunol.* **171**, 6251-9.

29. **N. LONBERG** (2005) Human antibodies from transgenic animals. *Nat Biotechnol.* **23**, 1117-25.

Awards and Grants:

National Science Foundation Predoctoral Fellowship,	1979-1982.
American Cancer Society Postdoctoral Fellowship,	1985-1988.
National Institutes of Health Postdoctoral Fellowship,	1988-1989.
National Institutes of Health SBIR Grant (phases I & II),	1991-1994.

Patents:

H.Meade and N. LONBERG (1989) Isolation of Exogenous Recombinant Proteins from the Milk of Transgenic Mammals. *U.S. Patent* #4,873,316.

N. LONBERG and R. Kay (1996) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies. *U.S. Patent* #5,545,806.

N. LONBERG and R. Kay (1996) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies of Various Isotypes. *U.S. Patent* #5,569,825.

N. LONBERG and R. Kay (1997) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies. *U.S. Patent* #5,625,126.

N. LONBERG and R. Kay (1997) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies. *U.S. Patent* #5,633,425.

N. LONBERG and R. Kay (1997) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies of Various Isotypes. *U.S. Patent* #5,661,016.

N. LONBERG and R. Kay (1997) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies. *European. Patent* #0 814 159 A2

H.Meade and N. LONBERG (1998) Transgenic Non-Human Mammal Milk. *U.S. Patent* #5,750,172.

N. LONBERG and R. Kay (1998) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies. *U.S. Patent* #5,770,429.

N. LONBERG and R. Kay (1998) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies. *U.S. Patent* #5,789,650.

N. LONBERG and R. Kay (1998) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies. *U.S. Patent* #5,814,318.

N. LONBERG and R. Kay (1999) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies. *U.S. Patent* #5,874,299.

N. LONBERG and R. Kay (1999) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies of Various Isotypes. *U.S. Patent* #5,877,397.

N. LONBERG and R. Kay (1999) Transgenic Non-Human Animals Capable of Producing Heterologous Antibodies. *Japanese. Patent* #2938569.

N. LONBERG and R. Kay (2001) High Affinity Human Antibodies and Human Antibodies Against Digoxin. *U.S. Patent* #6,255,458.

N. LONBERG and R. Kay (2001) Transgenic Non-Human Animals for Producing Heterologous Antibodies. *U.S. Patent* #6,300,129.

J. Buechler, G. Valkirs, J. Gray, and N. LONBERG (2004) Human Antibodies as Dignostic Reagents. *U.S. Patent* #6,680,209.

J. Buechler, G. Valkirs, J. Gray, and N. LONBERG (2004) Human Antibodies. *U.S. Patent* #6,794,132.

Dedicated to the memory of Georges Kohler

Immunoglobulin Genes
Second edition

Edited by

T. Honjo
Kyoto University, Japan

F. W. Alt
Howard Hughes Medical Institute Research Laboratories,
The Children's Hospital, Boston

ACADEMIC PRESS

Harcourt Brace & Company, Publishers
London, San Diego, New York, Boston, Sydney, Tokyo, Toronto

ACADEMIC PRESS LIMITED
24–28 Oval Road
LONDON NW1 7DX, UK

U. S. Edition Published by
ACADEMIC PRESS INC.
San Diego, CA 92101, USA

This book is printed on acid-free paper

A catalogue record for this book is available from the British Library

ISBN 0 12 053640 4

Typeset by Phoenix Photosetting, Chatham
Printed and bound by Hartnolls Ltd, Bodmin, Cornwall

Contents

Contributors

FW Alt
Howard Hughes Medical Institute Research Laboratories, The Childr Hospital, 300 Longwood Avenue, Boston, MA 02115, USA

MK Anderson
Department of Molecular Genetics, All Children's Hospital, 801 Sixth St South, St Petersburg, FL 33701-4899, USA

PD Burrows
Division of Developmental and Clinical Immunology, Departments of Mi biology, Medicine and Pediatrics, University of Alabama at Birmingham, Wallace Tumor Institute, UAB Station, Birmingham, AL, USA

K Calame
Columbia University College of Physicians and Surgeons, Departmer Microbiology, 701 West 168th Street, New York, NY 10032-2704, USA

JD Capra
Southwestern Medical School, Department of Microbiology, 5323 H Hines Boulevard, Dallas, TX 75235, USA

MD Cooper
Howard Hughes Medical Institute, Department of Pediatrics Microbiology, University of Alabama at Birmingham, 263 Wallace Tu Institute, UAB Station, Birmingham, AL 35294, USA

LE Daltch
Tufts University School of Medicine, Department of Pathology, 136 Harr Avenue, Boston, MA 02111, USA

EA Faust
Howard Hughes Medical Institute Research Laboratories, Universit California, Los Angeles, 5-748 MacDonald Building, 10833 Le C Avenue, Los Angeles, CA 90024-1662, USA

S Ghosh
Howard Hughes Medical Institute, Yale University, 295 Congress Ave 154 BCMM, New Haven, CT 06510, USA

LA Herzenberg
Stanford University School of Medicine, Department of Genetics, Imm genetics and Cell Sorting Laboratory, Stanford, CA 94305-1662, USA

8

The human immunoglobulin κ genes

Hans G. Zachau

Institut für Physiologische Chemie der Universität München, München, Germany

While the chapter on the immunoglobulin κ genes in the first edition of this book covered the information available in 1987 on the genes of human and mouse (Zachau, 1989a), the present review deals with the human κ genes only and concentrates on the more recent results. In fact, quotations to the older literature will not be repeated here. For the mouse $V_κ$ gene families, their complexity, polymorphism and use in non-autoimmune responses the reader is referred to a recent review by Kofler *et al.* (1992). Other aspects are dealt with in the respective chapters of this book.

THE ELUCIDATION OF THE HUMAN κ LOCUS

The single human $C_κ$ gene and the five $J_κ$ genes were cloned and characterized by P. Leder's group (Hieter *et al.*, 1980, 1982) and the first human $V_κ$ genes were isolated

IMMUNOGLOBULIN GENES (2ND EDITION)
ISBN 0-12-053640-4

and sequenced by Bentley and Rabbitts (1980, 1981, 1983). In our group, the κ genes have been studied since the early 1980s, and some aspects of the work were reviewed in lecture reports (Zachau, 1990, 1993) and in a recent survey (Zachau, 1995).

The structural work in our laboratory was based on 440 cosmid and 30 phage λ clones that were isolated from various libraries of germline DNA and mapped with 5–12 different restriction nucleases each. Initially, the search clones were $V_κ$ gene probes and specific chromosomal walking probes at later stages of the work. The first indications that there may be two copies of the κ locus came from an apparently duplicated $V_κ$ gene (Bentley and Rabbitts, 1983) and then from the systematic cloning studies of Pech *et al.* (1985). The extent of the duplication became known with some certainty only at a late state of the work. It is now clear that the so-called $C_κ$ proximal (p) copy of the locus contains in a 600-kb contig, in addition to the $J_κ$–$C_κ$ region, 40 $V_κ$ genes and pseudogenes, and the distal (d) copy contains in a 440-kb contig 36 $V_κ$ genes and pseudogenes. A scheme of the κ locus and its surroundings is presented for general orientation (Fig. 1). The two large contigs were assembled from smaller ones that had been studied separately before: Op/Od (Pargent *et al.*, 1991a), Ap/Ad (Lautner-Rieske *et al.*, 1992), Lp/Ld (Huber *et al.*, 1993a,b), B (Lorenz *et al.*, 1988) and $J_κ$–$C_κ$–κde (Klobeck and Zachau, 1986; Klobeck *et al.*, 1987a). An *Sna*BI map of the contigs was reported by Ermert (1994). The cloning of the gaps between these original

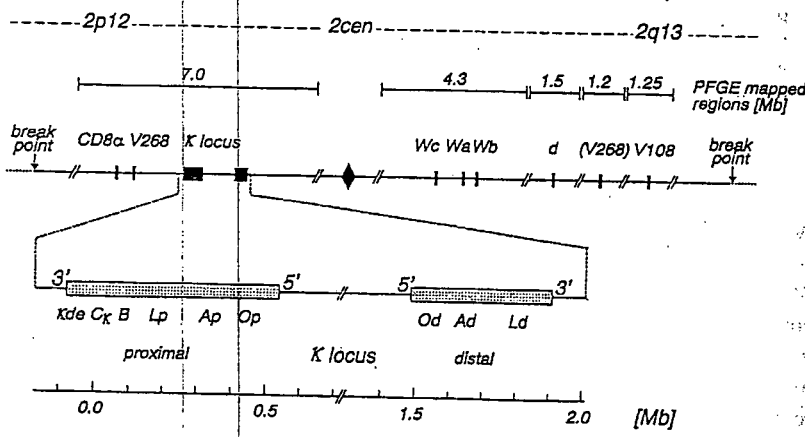


Fig. 1. Schematic representation of the central part of chromosome 2. The cloned regions of the κ locus are shown as black bars or stippled boxes. Vertical lines indicate cloned orphon regions (V268, Wa–Wc, V108; see p. 185) and fragments hybridizing to the following probes: CD8α and 273–2, derived from the V268 region and therefore designated (V268) (Weichhold *et al.*, 1993b; Huber *et al.*, 1994); d, homologous to the κde region (Graninger *et al.*, 1988). The organization of the κ locus is depicted according to Weichhold *et al.* (1993a). In the scale $C_κ$ is taken as zero, while for the contigs counting starts at the 5′ ends (Pargent *et al.*, 1991a). The orphons on the long arm and the breakpoints of pericentric inversion(s) are described by Lautner-Rieske *et al.* (1993).

small contigs by chromosomal walking was a lengthy and cumbersome process, since the clones required for linking were highly underrepresented in the libraries. Structural reasons for this did not become apparent when the linking was achieved.

THE $V_κ$ GENES OF THE LOCUS

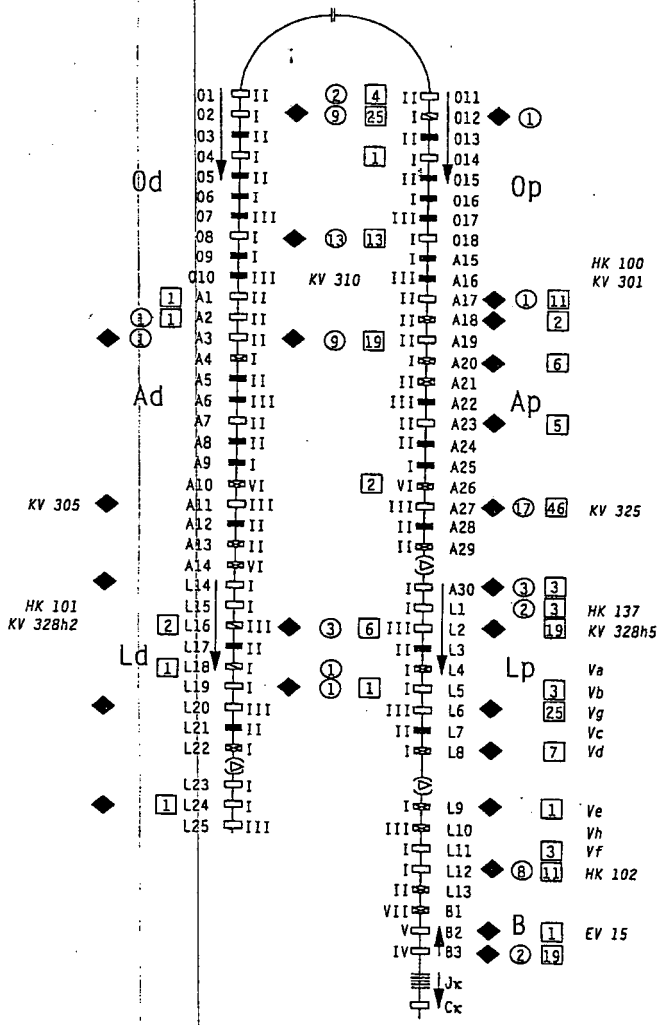
Genes and pseudogenes, subgroups, polymerase chain reaction (PCR) primers

The 76 $V_κ$ genes and pseudogenes of the locus were sequenced and their transcriptional polarities were determined within the maps. An outline of the $V_κ$ genes of the locus is given in Fig. 2. There are 10 solitary genes and 33 gene pairs whose sequences are 95–100% identical in the coding regions. Of the $V_κ$ genes 32 are potentially functional, 16 have minor defects and 25 are pseudogenes; three genes were found to occur both as potentially functional and slightly defective alleles. The minor defects are defined as one or two 1-bp alterations in a gene, for instance the occurrence of a stop codon and/or a deviation from the canonical sequences of regulatory elements, splice sites or hepta- and nona-nucleotide recognition sequences. The 16 genes with minor defects are defined as a separate class of genes, since, as for the three genes mentioned above, potentially functional alleles may exist in the human population. This is not to be expected for the pseudogenes, which usually carry several defects each. All human $V_κ$ gene, pseudogene and orphon sequences including all alleles known to us were compiled (Schäble and Zachau, 1993) and, in addition, some of the pseudogenes were dealt with specifically by Schäble *et al.* (1994). Recently, Cox *et al.* (1994) amplified from genomic DNA by PCR FR1–CDR3 sequences and called four of them 'new gene segments'. However, the sequences do not correspond to new gene loci but have to be considered alleles of published $V_κ$ genes and orphons, as far as this can be concluded in the absence of intron sequences and data on the genomic context (Klein and Zachau, 1995). The systematic nomenclature of $V_κ$ genes (Fig. 2) is used throughout this chapter and in all recent reports from our laboratory, but it is expected that some alternative designations will be further employed in the literature. The different designations of the various genes were compiled by Schäble and Zachau (1993).

There is good circumstantial evidence that we have now cloned all or most $V_κ$ genes of the locus. The reservations inevitably connected with such a statement have been discussed by Meindl *et al.* (1990a), Huber *et al.* (1993a,b) and Klein *et al.* (1993). One prerequisite was, of course, that we were able to close all gaps within the p and the d copies of the locus and to extend the two contigs in both directions by 50–80 kb without finding additional genes, an effort that is at present being continued with YAC clones (I. Zocher and J. Brensing-Küppers, unpublished data). Recently, one still Unidentified Hybridizing Object (UHO) was detected in one of the YAC clones (J. Brensing-Küppers, unpublished). However,

previously identified UHOs were either orphons (p. 184) or turned out on sequencing not to contain V_κ like structures but a LINE1 sequence and, in some cases, M13 vector sequences, the cross-hybridizations of which with human DNA are known (Vassart *et al.*, 1987); this can be taken as an indication that the search for additional V_κ genes by hybridization was carried to the limit (Röschenthaler *et al.*, 1992; Schäble *et al.*, 1994).

The classification of κ proteins into four subgroups (Kabat *et al.*, 1991 and earlier editions) was fully confirmed when the V_κ gene sequences were aligned (Schäble and Zachau, 1993). The similarity between potentially functional V_κ gene segments is



higher than 84% among the members of subgroups I–III and between 57 and 78% when members of different subgroups including pseudogenes are compared. While the one gene of subgroup IV is transcribed and translated, no proteins are known for the genes of subgroups V–VII. However, transcripts of V_κV and V_κVI genes have been found recently (Marks *et al.*, 1991). There are only five V_κ genes altogether in subgroups V–VII (Straubinger *et al.*, 1988a), but the definition of separate subgroups for them seems unavoidable if the members of a subgroup should be at least 80% similar to each other.

The alignment of the sequences of the V_κ gene regions (Schäble and Zachau, 1993) also served two other purposes. PCR primer combinations were derived, which should allow the reliable amplification of certain groups of germline genes and also some single genes. The other aim was to define and evaluate conserved sequence elements.

Conserved sequence elements

In the upstream region there is, in addition to the rather variable TATAA-like sequence, the decanucleotide (dc) sequence TNATTTGCAT, which was early recognized as a functional promoter (Falkner and Zachau, 1984). Independently, the octanucleotide sequence ATTTGCAT was defined as a conserved sequence (Parslow *et al.*, 1984). It is now seen in the alignments of the human V_κ gene sequences that the dc sequence is very largely conserved among potentially functional V_κ genes and that the heptanucleotide TTTGCAT is fully conserved. This is in line with the observation that these seven nucleotides are essential for promoter activity, while alterations in the first and third position of dc allow reduced transcription (Wirth *et al.*, 1987). A 15-mer or pd element (Falkner and Zachau, 1984) is found 17 bp 5′ of dc in all V_κI genes and about 150 bp 5′ of the V_κVI genes A10 and A26, while it is not seen in V_κ genes of the other subgroups. pd is not a promoter element essential for transcription (Bergman *et al.*, 1984) but it seems to have a supportive activity (Sigvardsson *et al.*, 1995). Another possibly

Fig. 2. Outline of the human κ locus. Open boxes represent potentially functional V_κ genes and the single C_κ gene, filled boxes the V_κ pseudogenes. Boxes with crossed lines designate genes with minor defects (as defined in the text) and boxes with one diagonal line the three genes for which potentially functional and slightly defective alleles are known. Roman numerals refer to the subgroups of the respective V_κ genes. The deletions in the A and L regions -(Δ)- are described on pp. 179 and 186. The drawing is not to scale. Arrows show the direction of transcription. The boxes, circles and rhomboids beside the V_κ genes or between undistinguishable gene pairs refer to transcription products, κ proteins and genomic joints, respectively. The figures in the boxes and circles are the numbers of different gene products found. Alternative designations of some V_κ genes used in the literature are shown in italics; a complete listing of such designations is given in Schäble and Zachau (1993). The figure is similar to previous published versions (Klein *et al.*, 1993; Klein and Zachau, 1995) and further details are described therein.

supportive element in the dc region is CCCT (Högbom *et al.*, 1991). An ACCC element nearby was found to bind nuclear proteins (Mocikat *et al.*, 1988). It is a matter of definition whether some of the elements, notably the dc element itself which is found in the opposite 5′, 3′ polarity upstream of V_H genes (Falkner and Zachau, 1984), serve a promoter or an enhancer function. The work on the upstream elements of κ genes was compiled by Mocikat *et al.* (1989) and a comprehensive review of the regulatory elements in immunoglobulin genes was given by Staudt and Lenardo (1991).

Another outcome of the V_κ alignments (Schäble and Zachau, 1993) is that within the major subgroups the leader segments are more similar to each other than the V gene segments. Some conserved regions in the introns are related to functions in the splice process. The intron sizes are remarkably similar within subgroups and show pronounced differences between subgroups. Deviations from the recombination signal sequences are compiled in Schäble *et al.* (1994).

THE B3–J_κ–C_κ-κde REGION

The 23-kb region between the single V_κIV gene of the locus, which is called B3, and J_κ1 was found to be free of V_κ gene-like structures in the DNAs of several individuals. However, a sequence of about 0.5 kb was found in the middle of the region that has a counterpart called homox on another chromosome but otherwise does not hybridize to genomic DNA (Klobeck *et al.*, 1989). No function is known for the two sequences, which are 96% identical. The finding of spliced J_κ–C_κ transcripts without a V_κ gene (Martin *et al.*, 1991) points to the existence of a promoter and a transcription start site about 4 kb upstream of J_κ1, which may function in a prelude to V_κ–J_κ rearrangements.

Although the sequences of the J_κ and C_κ genes have been known for a long time (Hieter *et al.*, 1980, 1982), the sequence of the whole J_κ–C_κ region of more than 5 kb has been made available only recently (Whitehurst *et al.*, 1992). This was done in the context of further defining the location of the matrix association region (MAR) in the J_κ–C_κ intron. Slightly upstream of MAR lies the κde target sequence (see below) and slightly downstream the intron enhancer, which was characterized by Gimble and Max (1987). The various enhancing and silencing sequence motifs were reviewed by Staudt and Lenardo (1991). An additional silencing element immediately upstream of the NF-κB binding site of the intron was recently described for the mouse and human systems (Saksela and Baltimore, 1993).

The C_κ allotypes, which were originally defined serologically and by protein sequencing, were now studied by PCR permitting easy detection of the association to other polymorphic markers of the region (Moxley and Gibbs, 1992). At a position 12 kb to the 3′ side of C_κ lies the so-called downstream enhancer, which was identified by sequence homology to the corresponding mouse enhancer (Müller *et al.*, 1990). Its functional characteristics were then defined by Judde and Max (1992).

Another 12 kb downstream lies the κde (C_κ deleting) element, which was recognized in the human system by Siminovitch *et al.* (1985) and localized by Klobeck and Zachau (1986). Its hepta- and nona-nucleotide recognition signals (located at 23 bp distance) recombine with the complementary signal sequences in the intron (located at 29 bp distance) leading to the excision of C_κ and the enhancers in some λ chain-producing B cells. Nothing is known yet about the enzymology and regulation of this process.

THE STRUCTURAL ORGANIZATION OF THE κ LOCUS IN GERMLINE DNA

The κ locus is located on the short arm of chromosome 2 at 2cen–p12 (Malcolm *et al.*, 1982; McBride *et al.*, 1982) or, more specifically, at 2cen–p11.2 (Lautner-Rieske *et al.*, 1993). Mapping by pulsed field gel electrophoresis (PFGE) with the help of 13 rare-cutter restriction nucleases and 15 unique hybridization probes led to a detailed picture of the locus and its surroundings (Weichhold *et al.*, 1993a). The p and d contigs are arranged in opposite 5′, 3′ polarity (Fig. 1). The still uncloned region of 800 kb between the contigs appears not to contain further V_κ genes (see p. 175). The structure is largely symmetrical starting from a centre in the uncloned region and extending for about 850 kb to each side, i.e. to the duplicate gene pair L10/L25 as depicted in Fig. 2. The map of the κ locus, comprising about 2 Mb, extends for another 1.5 Mb towards the centromere, but no marker is known between the locus and the centromere. The map of 3.5 Mb towards the telomere includes the orphon V_κ gene V268 (Huber *et al.*, 1994) and, at a distance of 2–2.2 Mb from C_κ, the CD8α locus (Weichhold *et al.*, 1993b). Some detours and artefacts in establishing the PFGE map were also described (Weichhold *et al.*, 1993c).

The 5′ termini of the p and d contigs, i.e. the regions towards the uncloned central part beyond the genes O1 and O10 (Fig. 2), are rich in repetitive DNA sequences (Pargent *et al.*, 1991a). The same seems to be the case at the 3′ end of the d contig beyond L25 (Huber *et al.*, 1993a). These sequences have possibly played a role in confining the κ locus to its present limits. The inverted duplicated structure of the κ locus may have been formed in reactions similar to those described in models of gene amplification (for a discussion see Weichhold *et al.*, 1993a).

Three regions in the cloned parts of the locus, i.e. between A29 and A30, L8 and L9, and L22 and L23 (Fig. 2), are not represented on the opposite copy. Since an artefactual loss of these regions on cloning can be excluded, they must have been deleted during evolution, albeit after the duplication of the locus. All deletion breakpoints were sequenced. The sequence motif CCAG/CTGG found by Chou and Morrison (1993) to occur commonly near (somatic) non-homologous recombination breakpoints involving immunoglobulin gene sequences was observed rather frequently in our sequences, but no accumulation near the breakpoints was seen.

POLYMORPHISMS IN THE κ LOCUS

Allelic differences and haplotypes

Although the V_κ gene probes of the major subgroups hybridize to many related genes and complicated patterns ensue, it has been possible to define some V_κ gene-related allelic polymorphisms (Turnbull *et al.*, 1987). One such polymorphism was linked to rheumatoid arthritis with a relative risk of 5 (Meindl *et al.*, 1990b). However, for systematic studies of polymorphisms single-copy probes are much preferred. The detection of allelic differences by restriction fragment length polymorphism (RFLP) studies is relatively straightforward in the non-duplicated part of the locus, as in the B3 region (Klobeck *et al.*, 1987a), the B3–J_κ intergenic region (Klobeck *et al.*, 1989) and in the C_κ region (Field *et al.*, 1987; Klobeck *et al.*, 1987a; Moxley and Gibbs, 1992; and earlier literature on C_κ allotypes). In the duplicated part of the locus most specific probes recognize the homologous parts of both copies. Although they are, therefore, not truely unique they are included for the present discussion in the group of single-copy probes. Five RFLPs were established in the duplicated O regions, which together with the three RFLPs of the B3–C_κ region served to define three basic and several derived haplotypes of the κ locus (Pargent *et al.*, 1991b).

There is little allelic variation in the gene regions: no variants at all were found in the C_κ genes of 50 unrelated individuals. 12 variants were identified in the B3 genes of 26 individuals, but all of them were located in the intron (Kurth and Cavalli-Sforza, 1994); the 1-bp difference between all their germline B3 gene sequences and the published sequence, which is pointed out by Kurth and Cavalli-Sforza (1994), results from aligning to the sequence of a rearranged and mutated B3 gene but not to the germline B3 gene sequence described in the same paper (Klobeck *et al.*, 1985). Schäble and Zachau (1993) compiled 22 alleles of 19 other V_κ genes. Of 27 different V_κ gene sequences with open reading frames reported by Cox *et al.* (1994) 26 were found to be identical to previously published sequences (that had been determined in the DNA of various individuals) and one had a 1-bp difference; an allele of a pseudogene pair and sequences related to two orphons were also reported. This is our interpretation of the respective data as described in Klein and Zachau (1995). The implications of allelic variation for mutation studies in V_κ genes are discussed on p. 183.

Duplication differentiating polymorphisms

If a hybridization probe recognizes homologous p- and d-copy derived fragments and the fragments are of different sizes, it defines a duplication differentiating polymorphism (DDP). With some DDP probes RFLPs were also detected but for the majority no allelic differences have been found as yet. The extent of duplication of the κ locus was determined with the help of 16 DDPs distributed over the entire locus (Pargent *et al.*, 1991b) and, of course, by the PFGE work (Lorenz *et al.*, 1987; Weichhold *et al.*, 1993a). The DDPs were essential in the structural work on the κ locus, since every newly isolated phage or cosmid clone had to be assigned to the p or d copy.

Haplotype 11

For most RFLPs and DDPs it is not known whether the underlying appearance or disappearance of a restriction site is caused by a base change, a deletion, insertion or by another structural change. One haplotype, however, is known to differ from the 'normal' haplotype N by the absence of the whole d copy of the locus. This so-called haplotype 11 was found in an individual homozygous for it (Straubinger *et al.*, 1988b). The haplotype was characterized by hybridization to DDP probes across the locus (Pargent *et al.*, 1991b). In a group of 23 caucasoid individuals there was, in addition to the homozygous one, one heterozygous individual. In a group including individuals of African and Asian origin 2 of 41 individuals were found to be heterozygous for haplotype 11 (Schaible *et al.*, 1993). In PFGE experiments it was shown that about 1.0 Mb, including the whole d contig, is absent from the DNA of the homozygous individual, and indirect evidence indicates that this is due to a deletion rather than to the persistence of an evolutionarily early non-duplicated state (Weichhold *et al.*, 1993a).

REARRANGED V_κ GENES

Since this topic is also dealt with in other chapters of this book, only some aspects related to the structure of the human κ locus are covered here. The available data on the mechanism of V(D)J joining and on V–J, V–D and DJ junctions were comprehensively reviewed by Lewis (1994). Among the reviews on hypermutation the recent ones by Berek (1993) MacLennan (1994) and Hengstchläger *et al.* (1995) should be mentioned.

V_κ–J_κ rearrangements

The 5′, 3′–polarity of the V_κ genes within the locus (arrows in Fig. 2) determines the type of rearrangement: the two J_κ–proximal genes B2 and B3, whose polarity is opposite to that of the J_κ–C_κ segment, are rearranged by an inversion mechanism (Klobeck *et al.*, 1987a; Lorenz *et al.*, 1988), while the other V_κ genes of the p copy rearrange by deletion of the stretch of DNA between the V_κ and J_κ genes (Weichhold *et al.*, 1990). The genes of the d copy are located 1.35–1.8 Mb from J_κ–C_κ and the polarities of all of them are opposite to the polarity of J_κ–C_κ. For one of the d-copy genes the rearrangment by inversion was proven by PFGE experiments (Weichhold

et al., 1990). Since for all p- and d-copy genes the polarities were determined by sequencing and detailed restriction mapping, their mode of rearrangement can be inferred from the cases that had been studied in detail.

The reciprocal products to the V_κ–J_κ joints are the signal joints, in which the hepta- and nona-nucleotide recombination sequences are linked back to back. The first signal joint was found in genomic DNA by Steinmetz *et al.* (1980) and was interpreted by Lewis *et al.* (1982) as the product of an inversion. Several signal joints have been found in the human κ locus (reviewed in Zachau, 1989a and Klein *et al.*, 1993). When V and J gene segments are joined by the deletion mechanism, the excised material is lost from the cells or found as circular DNA. Such circles also carrying signal joints have been found in several systems including the mouse κ system (e.g. Hirama *et al.*, 1991) but, possibly for technical reasons, not yet for the human κ genes.

B cells can undergo consecutive V_κ–J_κ rearrangements until all J_κ elements have been used. Examples of an inversion followed by a deletion and of two subsequent inversions have been reported (e.g. Klobeck *et al.*, 1987a; Lorenz *et al.*, 1988). In one cell line all products of a deletional V_κ–J_κ joining (in combination with a t(2;8) translocation; Klobeck *et al.*, 1987b) and two consecutive inversions were cloned and sequenced. In the second rearrangement a productive V_κ–J_κ–C_κ joint was produced and, contrary to the common assumptions, this did not prevent a further recombination, which in this case was an aberrant one (Huber *et al.*, 1992).

Apparently, the recombination machinery can handle inversions of 25–42-kb fragments for the genes B3 and B2, and of megabase-sized, that is millimetre-long, fragments for the d-copy genes. An intermediate formation of looped chromatin threads would have to be assumed. Clearly any deletional V_κ–J_κ joining leads to a loss of V_κ genes from the genome, while in an inversional joining all V_κ genes stay in the genome and can, in principle, be used in a second round of recombination. Also, genes other than V_κ, whose existence in the 800 kb between the p and d copies cannot be excluded (see below), would be kept in the genome on the (inversional) rearrangement of d-copy genes.

Which V_κ genes of the locus are rearranged, transcribed and translated?

This question was addressed by Klein *et al.* (1993) and Klein and Zachau (1995) on the basis of 70 of our own cDNA sequences and numerous nucleic acid and protein sequences from the literature. The results are shown in Fig. 2. It can be seen that some germline genes give rise to many transcripts and proteins and others to much fewer ones. For eight genes that, by definition (see p. 175), are potentially functional no transcripts or proteins were found. In general, fewer products were found to be derived from the d-copy than from the p-copy genes, although some cDNAs or proteins, whose sequences fit both duplicated germline genes (middle section in Fig. 2), may well be derived from d-copy genes. Figure 2 represents our current state of

knowledge. Other genomic V_κ–J_κ joints and κ proteins may still be found, since the currently known ones are not the outcome of systematic searches. Transcription products (cDNAs), on the other hand, have been screened for in several laboratories. However, here also the absence of products for a potentially functional gene does not necessarily reflect its inability to be rearranged and transcribed, since the cDNA libraries are, of course, the outcome of immunological selection in the particular B cell repertoires under study; also in some cases an experimental bias in screening of the libraries cannot be excluded. At present, of the 76 V_κ genes of the locus 22 genes and five pairs of duplicated identical genes are known to be transcribed. The corresponding numbers for rearranged genomic V_κ genes and for full-length κ proteins are 17 plus 4 and 7 plus 7.

Somatic mutation

In considering the extent and type of somatic mutations, the definite assignment of the rearranged V_κ genes to certain germline genes is essential. The first mutated V_κ genes found (in the mouse system) could be defined, because the genomic surroundings of the rearranged and the unrearranged genes helped with the assignments (Pech *et al.*, 1981). The assignment of cDNAs to germline genes is more difficult. In the human system this became possible only after it could be reasonably assumed that all functional germline genes of the locus were known. Because of allelic variation, the sequence of the unrearranged germline gene of the same individual should be known from whom the mutated V_κ–J_κ gene or cDNA is derived (see p. 180). This demand has to be taken seriously if one is interested in the mutation behaviour of specific single V_κ genes. It is less important if one considers the average extent and type of mutations of large numbers of rearranged genes or cDNAs.

There is a wide range of numbers of mutations per gene: for instance a cDNA with no mutation was isolated from the same library as other cDNAs with 25 mutations, which were derived from the same germline gene pair (Klein *et al.*, 1993). The survey of the cDNA sequences from our laboratory and the data from the literature gave information on the type of mutations and on various other features of the mutation process, but only one feature should be mentioned here — the fact that mutations in adjacent nucleotides are found about twice as often as expected statistically. On a nucleotide basis, 20–25% of the mutations are in such blocks. About half of our cDNA clones and about 40% of the human V_κ sequences from the literature carry block mutations. They may have arisen from independent mutations in adjacent nucleotides and/or from combined exchange processes. The occurrence of blocks of altered nucleotides may be a feature of the maturation of the immune response in human and mouse. Certainly, block mutations that lead to amino acid replacements would be subject to selection. Somatic gene conversion appears not to play a major role in introducing single or block mutations in the human κ system. The mechanistic features possibly involved in the hypermutation and block mutation processes of V_κ genes were discussed by Klein *et al.* (1993), Klein and Zachau (1995) and Zachau (1995).

Sequences within or downstream of the J_κ–C_κ region are probably important for the hypermutation process since in an aberrantly rearranged κ gene, which is broken in V_κ by a t(2;8) translocation, the mutations extend to the adjacent chromosome 8 sequences but are not found in the 5′ part of the V_κ gene (Klobeck *et al.*, 1987b). A more detailed study in the mouse system was reported by Betz *et al.* (1994). Somatic mutations in an aberrantly rearranged V_κ gene had been found previously (Pech *et al.*, 1981).

V_κ–J_κ junctions

There is an accumulation of base changes in the V_κ gene sequences close to the junction that is caused, at least in part, by the truncation and repair processes found in many V(D)J joining systems (review by Lewis, 1994). Therefore, the assignment of the V-gene moiety of the junction to a certain germline gene has to be based on a comparison of full-length sequences. In the survey of cDNA sequences from our laboratory and of data from the literature (Klein *et al.*, 1993; Klein and Zachau, 1995) about one-fifth of the V_κ–J_κ junctions contained additional nucleotides between V_κ and J_κ. These nucleotides code for amino acids 95A and 95B or, if their number is different from 3 or 6, destroy the reading frame. The additional nucleotides do not have the characteristics of P or palindromic elements (Lafaille *et al.*, 1989; Roth *et al.*, 1992). They may be derived from the germline nucleotides between the last canonical codon, i.e. codon 95, and the heptanucleotide on the one side and between the complementary heptanucleotide and the first J_κ codon on the other side. The sequences of the additional nucleotides fit those germline sequences in several cases either fully or with one base change, which would have to be attributed to somatic mutations. Although terminal deoxynucleotidyl transferase (TdT) is generally not detected in B cells at the time the κ genes are rearranged, the presence of N segments (Alt and Baltimore, 1982) in these genes has been reported repeatedly (Klobeck *et al.*, 1987b; Martin *et al.*, 1991; H. Schroeder, personal communication). About 80% of the additional nucleotides in V_κ–J_κ junctions are C and G residues. This fits what would be expected for TdT-catalysed insertions, but it also fits the composition of germline nucleotides adjacent to those V_κ and J_κ genes that are frequently found in V_κ–J_κ joints (Klein *et al.*, 1993; Victor and Capra, 1994). Therefore, it has to be checked in every case whether it is more likely that the additional nucleotides are germline or TdT derived or inserted by still another mechanism.

DISPERSED V_κ GENES

V_κ genes that are located outside the κ locus were called orphons in analogy to the histone and ribosomal RNA genes found outside the respective loci (Childs *et al.*, 1981). The V_κ orphons were discovered when it proved impossible to link by

chromosomal walking certain V_κ gene-containing cosmid clones to the existing contigs of the κ locus (Lötscher *et al.*, 1986). The true locations of the orphon V_κ clones were first shown with the help of panels of human–rodent cell hybrid DNAs and later by *in situ* hybridization. Twenty-four orphon V_κ genes have been cloned and sequenced. One of them is localized on chromosome 1 and a cluster of five V_κ genes on chromosome 22 (Lötscher *et al.*, 1986, 1988a). For five V_κI orphons it is only known that they are very similar to each other in sequence but not identical and that they are located on chromosomes other than chromosome 2 (Straubinger *et al.*, 1988c; Röschenthaler *et al.*, 1992). This so-called Z family of V_κI orphons may, in fact, have several more members (Meindl *et al.*, 1990a). Two yeast artificial chromosome (YAC) clones and three cosmid clones with restriction maps and hybridization properties similar to those of Z-orphon clones were isolated but not studied in detail (Pargent, 1991; Huber, 1993). Because of the high sequence similarity between the Z orphons it would require much effort to specify whether newly isolated clones are derived from independent loci or whether they are alleles of already known orphons (Röschenthaler *et al.*, 1992).

Of the 24 sequenced orphons 13 are localized on chromosome 2. One of them is located 1.5 Mb 3′ of C_κ (V268 in Fig. 1). According to its sequence this V_κ gene is potentially functional (Huber *et al.*, 1994) and a V_κ–J_κ rearrangement by inversion would not involve larger fragments than the rearrangements of V_κ genes of the d copy of the locus. However, no rearrangement products have been found yet. Because of its location outside the locus the gene is classified as an orphon. Another V_κ orphon without sequence defects is located on the long arm of chromosome 2 (V108 in Fig. 1; Huber *et al.*, 1990). The 11 V_κ orphons of the W regions (Fig. 1), on the other hand, are pseudogenes also according to their sequences (Zimmer *et al.*, 1990a). The three groups of W orphons were mapped to a 4.3-Mb region (Fig. 1; Weichhold *et al.*, 1992). They were probably derived from gene regions of the κ locus by a pericentric inversion and subsequent amplification events (Zimmer *et al.*, 1990b).

All V_κ orphons contain introns and, therefore, should have been dispersed on the DNA and not on the RNA and retrotranscript level; also, germline V_κ genes are believed not to be transcribed. Since, according to their sequences, all orphons, with one exception, have their closest relatives in the O regions (Schäble *et al.*, 1994), they may have been derived in evolution from a common precursor. However, the structural similarities between orphons and κ locus regions are not high enough to allow a duplicative mechanism of orphon formation to be postulated. The attempts to specify other features of the dispersion mechanism(s) have not been very successful. One such feature is the presence of sequences in the neighbourhood of some orphons that were supposed to bind replication and/or transcription factors (Lötscher *et al.*, 1988b), but no convincing arguments could be derived from that. At break-off points of homology between different orphon regions, which are probably junctions between translocated and receiving structures, direct and inverted repeats and an *Alu* element were found (Borden *et al.*, 1990). However, since the sequence features differed from one insertion break-point to the other, no unique mechanism of translocation could be proposed.

REPETITIVE AND UNIQUE SEQUENCES IN THE κ LOCUS

The κ locus was not specifically investigated for repetitive elements but 15 LINE1 and 25 *Alu* sequences were detected in hybridization experiments and/or sequence comparisons. The properties of the elements and some evolutionary considerations were compiled by Schäble *et al.* (1994). In the same report the unique sequences, which qualify as sequence-tagged sites (STS) as defined by Olson *et al.* (1989), are described. Such sequences are an important feature of the Human Genome Project, since they should allow the reproducible detection via PCR of certain chromosomal sites or the isolation of the respective clones from libraries. The STS sequences are distributed fairly well across the κ locus.

EVOLUTION OF THE VK GENES

A crucial event in the recent history of the κ locus was its duplication. The sequenced regions of the p- and d-copy genes and pseudogenes differ on average by about 1% (404 of 38 136 bp; Schäble and Zachau, 1993). If one assumes 1% of divergence to correspond to 1 million years of evolution (Wilson *et al.*, 1987), this should be the age of the duplicated locus. However, there are various caveats. First, the extent of divergence is highly uneven across the locus ranging from 0 to 3.7% for different gene regions, which may be interpreted in terms of a surveillance mechanism counteracting in certain regions the mutational divergence. In addition, the basic relation of 1% mutation per 10^6 years is certainly not undisputed. However, the postulated date of duplication may be roughly right since the κ locus of the chimpanzee seems not to be duplicated (Ermert *et al.*, 1995) and the time point in evolution when human and chimpanzee clades diverged may have been 4–5 million years ago. V_κ and C_κ sequences of non-human primates are very similar to human ones, e.g. the C_κ sequences of the chimpanzee and human are 99.6% identical.

Many events in the evolution of the κ locus occurred long before its duplication, e.g. the interdigitation of V_κ genes of different subgroups (Pech and Zachau, 1984; review Zachau, 1989b) and the duplication of a group of three V_κ genes (Huber *et al.*, 1993a). Also most changes, which converted V_κ genes to pseudogenes, happened before the duplication of the locus. Another old feature of the κ locus is the dispersion of V_κ genes to other parts of the genome. Since cosmid clones from the orphon regions of chromosomes 1 and 22 hybridized *in situ* to the assumed homologous chromosome bands of all great apes (Arnold *et al.*, 1995), the translocations may have happened very early in primate evolution. In the same study, a cosmid clone from one of the W regions was found to hybridize to a site that was pericentrically inverted in the corresponding chromosome of the chimpanzee, but not in that of the gorilla. Accordingly, the transposition occurred after the gorilla and before the chimpanzee clades diverged from the human evolutionary tree. This is also the time when the V108 region (see above; Fig. 1) became dissociated from the κ locus (Ermert *et al.*, 1995). The amplification of the transposed W region then occurred in at least two steps, which are postulated to have taken place 2×10^6 and 10^5 years ago, respectively (Zimmer *et al.*, 1990a).

Events that have to be dated after the duplication of the κ locus are the deletions of parts of Ap, Ld and Lp (Lautner-Rieske *et al.*, 1992; Huber *et al.*, 1993a), the insertion of an *Alu* element into one but not the other copy of the locus (Lautner-Rieske *et al.*, 1992; Schäble *et al.*, 1994) and at least some of the gene conversion-like events in the L regions (Huber *et al.*, 1993b). Most events that led to the divergence of the copies of the κ locus were point mutations. Not surprisingly, twice as many transitions as transversions are observed (Schäble and Zachau, 1993).

BIOMEDICAL IMPLICATIONS

Since there is reasonable certainty by now that all functional germline V_κ genes of the locus are known, conclusions are possible as to which part of the repertoire is expressed at which time in the development of the immune response. The κ chains found in pathological conditions, such as as autoimmune diseases or lymphomas, can be assigned to certain germline V_κ genes (compiled in Klein *et al.*, 1993; Klein *et al.*, 1995).

The individual who lacks, in a homozygous fashion, the d copy of the κ locus with its 36 V_κ genes (individual and haplotype 11; see p. 181) is apparently healthy and his κ chain/λ chain ratio is not altered (Schäble and Zachau, 1993). In general, the d-copy genes are not expressed to a great extent (Fig. 2) but it is known that the d-copy gene A2 codes for the most common light chain in the *Haemophilus influenzae* response (Scott *et al.*, 1989). Vaccination of individual 11 with the appropriate carbohydrate vaccine gave rise to antibodies whose light chains were derived, of course, from p-copy genes, but these light chains contained more somatic mutations than the usual A2-derived light chains (Scott *et al.*, 1991; review Scott *et al.*, 1992).

There is a 0.1% incidence of pericentric inversions of chromosome 2 in the present-day population. They have been formed in a process that is supposedly still going on. Heterozygous and homozygous (Gelman-Kohan *et al.*, 1993) individuals are apparently healthy. The inverted chromosomes carry the κ locus on the long arm and the W orphons on the short arm. The breakpoint on the long arm is, at the present level of analysis, indistinguishable from the one of the pericentric inversion that occurred in evolution (Lautner-Rieske *et al.*, 1993; previous literature quoted therein).

MISCELLANEOUS AND CONCLUDING REMARKS

The κ locus and its immediate surroundings comprise 2–3 Mb (Fig. 1), i.e. somewhat less than 0.1% of the human genome; 1040 kb of the locus and 800 kb of κ-related

sequences from outside the locus have been cloned and mapped at high resolution; 160 kb and 90 kb respectively of the clones have been sequenced, mostly gene regions, deletion breakpoints and the like. An additional 11 Mb in the neighbourhood of the κ locus and the orphons have been mapped at medium or low resolution. Further work on the κ locus, particularly large-scale sequencing studies, will contribute to the general understanding of genome structure and evolution, and there will always be surprising results in this type of work. One of the more interesting questions is whether there are non-V_κ open reading frames between the known V_κ genes and particularly in the still uncloned part between the p and d copies of the locus. The related question of detecting non-V_κ transcripts from the germline κ locus was studied recently and the results were clearly negative for a number of cell lines (Lautner-Rieske *et al.*, 1995). In the gap between the p and d copies clusters of rare-cutter restriction sites were observed (Weichhold *et al.*, 1993a) as they are found in CpG islands adjacent to housekeeping genes. However, only a combination of further cloning, sequencing and expression studies can reveal whether there are non-V_κ genes located within or near the κ locus.

The study of the κ genes has advanced to a state where the germline and expressed repertoires are largely known. Some open questions are general problems of immunogenetics, such as the enzymology and mechanisms of V(D)J joining, V-gene maturation by hypermutation and selection, and the switch from κ to λ gene expression. The known structures provide a basis for further mechanistic studies aimed at answering these questions.

NOTE ADDED IN PROOF

This review was concluded at the time of submission in March 1994 and only partially updated in the proofs in June 1995.

Acknowledgements

I thank the members of our group for their contributions and for the discussion of the present manuscript. The work of our laboratory was supported by the Bundesministerium für Forschung und Technologie, Center Grant 0316200A, and the Fonds der Chemischen Industrie.

REFERENCES

Alt, F.W. and Baltimore, D. (1982). *Proc. Natl Acad. Sci. USA.* **79**, 4118–4122.
Arnold, N., Wienberg, J., Ermert, K. and Zachau, H.G. (1995). *Genomics* **26**, 147–150.
Bentley, D.L. and Rabbitts, T.H. (1980). *Nature* **288**, 730–733.
Bentley, D.L. and Rabbitts, T.H. (1981). *Cell* **24**, 613–623.
Bentley, D.L. and Rabbitts, T.H. (1983). *Cell* **32**, 181–189.

Berek, C. (1993). *Curr. Opin. Immunol.* **5**, 218–222.
Bergman, Y., Rice, D., Grosschedl, R. and Baltimore, D. (1984). *Proc. Natl Acad. Sci. USA.* **81**, 7041–7045.
Betz, A.G., Milstein, C., González-Fernandez, A., Pannell, R., Larson, T. and Neuberger, M.S. (1994). *Cell* **77**, 239–248.
Borden, P., Jaenichen, R. and Zachau, H.G. (1990). *Nucleic Acids Res.* **18**, 2101–2107.
Childs, G., Maxson, R., Cohn, R.H. and Kedes, L. (1981). Cell **23**, 651–663.
Chou, C.L. and Morrison, S.L. (1993). *J. Immunol.* **150**, 5350–5360.
Cox, J.P.L., Tomlinson, I.M. and Winter, G. (1994). *Eur. J. Immunol.* **24**, 827–836.
Ermert, K. (1994). Doctoral Thesis, Fakultät für Chemie und Pharmazie, Universität München.
Ermert, K., Mitlöhner, H., Schempp, W. and Zachau, H.G. (1995). *Genomics* **25**, 623–629.
Falkner, F.G. and Zachau, H.G. (1984). *Nature* **310**, 71–74.
Field, L.L., Tobias, R. and Bech-Hansen, T. (1987). *Nucleic Acids Res.* **15**, 3942.
Gelman-Kohan, Z., Rosensaft, J., Nisani Ben-Cohen, R. and Chemke, J. (1993). *Hum. Genet.* **92**, 427.
Gimble, J.M. and Max, E.E. (1987). *Mol. Cell. Biol.* **7**, 15–25.
Grüninger, W.B., Goldman, P.L., Morton, C.C., O'Brien, S.J. and Korsmeyer, S.J. (1988). *J. Exp. Med.* **167**, 488–501.
Hengstschläger, M., Maizels, N. and Leung, H. (1995) In 'Progress in Nucleic Acid Research and Molecular Biology' (W.E. Cohn and K. Moldave, eds.) Vol. 50, pp. 67–99. Academic Press, New York.
Hieter, P.A., Max, E.E., Seidman, J.G., Maizel, J.V. and Leder, P. (1980). *Cell* **22**, 197–207.
Hieter, P.A., Maizel, J.V. and Leder, P. (1982). *J. Biol. Chem.* **257**, 1516–1522.
Hirama, T., Takeshita, S., Yoshida, Y. and Yamagishi, H. (1991). *Immunol. Lett.* **27**, 19–24.
Högbom, E., Magnusson, A.-C. and Leanderson, T. (1991). *Nucleic Acids Res.* **19**, 4347–4354.
Huber, C. (1993). Doctoral Thesis, Fakultät für Chemie und Pharmazie, Universität München.
Huber, C., Thiebe, R., Hameister, H., Smola, H., Lötscher, E. and Zachau, H.G. (1990). *Nucleic Acids Res.* **18**, 3475–3478.
Huber, C., Klobeck, H.-G. and Zachau, H.G. (1992). *Eur. J. Immunol.* **22**, 1561–1565.
Huber, C., Huber, E., Lautner-Rieske, A., Schäble, K.F. and Zachau, H.G. (1993a). *Eur. J. Immunol.* **23**, 2860–2867.
Huber, C., Schäble, K.F., Huber, E., Klein, R., Meindl, A., Thiebe, R., Lamm, R. and Zachau, H.G. (1993b). *Eur. J. Immunol.* **23**, 2868–2875.
Huber, C., Thiebe, R. and Zachau H.G. (1994). *Genomics* **22**, 213–215.
Judde, J.-G. and Max, E.E. (1992). *Mol. Cell. Biol.* **12**, 5206–5216.
Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S. and Foeller, C. (1991). 'Sequences of Proteins of Immunological Interest.' NIH Publication No. 91–3242.
Klein, R. and Zachau, H.G. (1995). *Ann. N.Y. Acad. Sci.*, in press.
Klein, R., Jaenichen, R. and Zachau, H.G. (1993). *Eur. J. Immunol.* **23**, 3248–3271.
Klein, U., Klein, G., Ehlin-Henriksson, B., Rajewsky, K. and Küppers, R. (1995). *Molecular Medicine*, in press.
Klobeck, H.-G. and Zachau, H.G. (1986). *Nucleic Acids Res.* **14**, 4591–4603.
Klobeck, H.-G., Bornkamm, G.W., Combriato, G., Mocikat, R., Pohlenz, H.D. and Zachau, H.G. (1985). *Nucleic Acids Res.* **13**, 6515–6529.
Klobeck, H.-G., Zimmer, F.-J., Combriato, G. and Zachau, H.G. (1987a). *Nucleic Acids Res.* **15**, 9655–9665.
Klobeck, H.-G., Combriato, G. and Zachau, H.G. (1987b). *Nucleic Acids Res.* **15**, 4877–4888.
Klobeck, H.-G., Combriato, G. and Zachau, H.G. (1989). *Biol. Chem. Hoppe-Seyler* **370**, 1007–1012.
Kofler, R., Geley, S., Kofler, H. and Helmberg, A. (1992). *Immunol. Rev.* **128**, 5–21.
Kurth, J.H. and Cavalli-Sforza, L.L. (1994). *Am. J. Hum. Genet.* **54**, 1037–1041.
Lafaille, J.J., DeCloux, A., Bonneville, M., Takagaki, Y. and Tonegawa, S., (1989). *Cell* **59**, 859–870.
Lautner-Rieske, A., Thiebe, R. and Zachau, H.G. (1995). *Gene*, **159**, 199–202.
Lautner-Rieske, A., Huber, C., Meindl, A., Pargent, W., Schäble, K.F., Thiebe, R., Zocher, I. and Zachau, H.G. (1992). *Eur. J. Immunol.* **22**, 1023–1029.
Lautner-Rieske, A., Hameister, H., Barbi G. and Zachau, H.G. (1993). *Genomics* **16**, 497–502.

Lewis, S., Rosenberg, N., Alt, F. and Baltimore, D. (1982). *Cell* **30**, 807–816.
Lewis, S.M. (1994). *Adv. Immunol.* **56**, 27–150.
Lorenz, W., Straubinger, B. and Zachau, H.G. (1987). *Nucleic Acids Res.* **15**, 9667–9676.
Lorenz, W., Schäble, K.F., Thiebe, R., Stavnezer, J. and Zachau, H.G. (1988). *Mol. Immunol.* **25**, 479–484.
Lötscher, E., Grzeschik, K.-H., Bauer, H.-G., Pohlenz, H.-D., Straubinger, B. and Zachau, H.G. (1986). *Nature* **320**, 456–458.
Lötscher, E., Zimmer, F.-J., Klopstock, T., Grzeschik, K.-H., Jaenichen, R., Straubinger, B. and Zachau, H.G. (1988a). *Gene* **69**, 215–223.
Lötscher, E., Siwka, W., Zimmer, F.-J., Grummt, F. and Zachau, H.G. (1988b). *Gene* **69**, 225–236.
McBride, O.W., Hieter, P.A., Hollis, G.F., Swan, D., Otey, M.G. and Leder, P. (1982). *J. Exp. Med.* **155**, 1480–1490.
MacLennan, I.C.M. (1994). *Cur. Biol.* **4**, 70–72.
Malcolm, S., Barton, P., Murphy, C., Ferguson-Smith, M.A., Bentley, D.L. and Rabbitts, T.H. (1982). *Proc. Natl Acad. Sci. USA* **79**, 4957–4961.
Marks, J.D., Tristem, M., Karpas, A. and Winter G. (1991). *J. Immunol.* **21**, 985–991.
Martin, D., Huang, R., LeBien, T. and Van Ness, B. (1991). *J. Exp. Med.* **173**, 639–645.
Meindl, A., Klobeck, H.-G., Ohnheiser, R. and Zachau, H.G. (1990a). *Eur. J. Immunol.* **20**, 1855–1863.
Meindl, A., Kellner, B., Schattenkirchner, M. and Zachau, H.G. (1990b). *Exp. Clin. Immunogenet.* **7**, 20–25.
Mocikat, R., Pruijn, G.J.M., van der Vliet, P.E. and Zachau, H.G. (1988). *Nucleic Acids Res.* **16**, 3693–3704.
Mocikat, R., Falkner, F.-G. and Zachau, H.G. (1989). In 'Tissue Specific Gene Expression' (R. Renkawitz ed.), pp. 73–85. VCH Verlag, Weinheim.
Moxley, G. and Gibbs, R.S. (1992). *Genomics* **13**, 104–108.
Müller, B., Stappert, H. and Reth, M. (1990). *Eur. J. Immunol.* **20**, 1409–1411.
Olson, M., Hood, L., Cantor, C. and Botstein, D. (1989). *Science* **245**, 1434–1435.
Pargent, W. (1991). Doctoral Thesis, Fakultät für Biologie, Universität München.
Pargent, W., Meindl, A., Thiebe, R., Mitzel, S. and Zachau, H.G. (1991a). *Eur. J. Immunol.* **21**, 1821–1827
Pargent, W., Schäble, K.F. and Zachau, H.G. (1991b). *Eur. J. Immunol.* **21**, 1829–1835.
Parslow, T.G., Blair, D.L., Murphy, W.J. and Granner, D.K. (1984). *Proc. Natl Acad. Sci. USA* **81**, 2650–2654.
Pech, M. and Zachau, H.G. (1984). *Nucleic Acids Res.* **12**, 9229–9236.
Pech, M., Höchtl, J., Schnell, H. and Zachau, H.G. (1981). *Nature* **291**, 668–670.
Pech, M., Smola, H., Pohlenz, H.-D., Straubinger, B., Gerl, R. and Zachau, H.G. (1985). *J. Mol. Biol.* **183**, 291–299.
Röschenthaler, F., Schäble, K.F., Thiebe, R. and Zachau, H.G. (1992). *Biol. Chem. Hoppe-Seyler* **373**, 177–186.
Roth, D.B., Menetski, J.P., Nakajima, P.B., Bosma, M.J. and Gellert, M. (1992). *Cell* **70**, 983–991.
Saksela, K. and Baltimore, D. (1993). *Mol. Cell. Biol.* **13**, 3698–3705.
Schäble, K.F. and Zachau, H.G. (1993). *Biol. Chem. Hoppe-Seyler* **374**, 1001–1022.
Schäble, K.F., Thiebe, R., Meindl, A., Flügel, A. and Zachau, H.G. (1994). *Biol. Chem. Hoppe-Seyler* **375**, 189–199.
Schaible, G., Rappold, G.A., Pargent, W. and Zachau, H.G. (1993). *Hum. Genet.* **91**, 261–267; Erratum **92**, 195.
Scott, M.G., Crimmins, D.L., McCourt, D.W., Zocher, I., Thiebe, R., Zachau, H.G. and Nahm, M.H. (1989). *J. Immunol.* **143**, 4110–4116.
Scott, M.G, Crimmins, D.L., McCourt, D.W., Chung, G., Schäble, K.F., Thiebe, R., Quenzel, E.-M., Zachau, H.G. and Nahm, M.H. (1991). *J. Immunol.* **147**, 4007–4013.
Scott, M.G., Zachau, H.G. and Nahm, M.H. (1992). *Int. Rev. Immunol.* **9**, 45–55.
Sigvardsson, M., Bemark, M. and Leanderson, T. (1995). *Mol. Cell Biol.* **15**, 1343–1352.
Siminovitch, K.A., Bakhshi, A., Goldman, P. and Korsmeyer, S.J. (1985). *Nature* **316**, 260–262.
Staudt, L.M. and Lenardo, M.J. (1991). *Annu. Rev. Immunol.* **9**, 373–398.
Steinmetz, M., Altenburger, W. and Zachau, H.G. (1980). *Nucleic Acids Res.* **8**, 1709–1720.

Straubinger, B., Thiebe, R., Huber, C., Osterholzer, E. and Zachau, H.G. (1988a). *Biol. Chem. Hoppe-Seyler* **369**, 601–607.
Straubinger, B., Huber, E., Lorenz, W., Osterholzer, E., Pargent, W., Pech, M., Pohlenz, H.-D., Zimmer, F.-J. and Zachau, H.G. (1988b). *J. Mol. Biol.* **199**, 23–34.
Straubinger, B., Thiebe, R., Pech, M. and Zachau, H.G. (1988c). *Gene* **69**, 209–214.
Turnbull, I.F., Sriprakash, K.S. and Mathews, J.D. (1987). *Immunogenetics* **25**, 193–199.
Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A.S. and Christophe, D. (1987). *Science* **235**, 683–684.
Victor, K.D. and Capra, J.D. (1994). *Mol. Immunol.* **31**, 39–46.
Weichhold, G.M., Klobeck, H.-G., Ohnheiser, R., Combriato, G. and Zachau, H.G. (1990). *Nature* **347**, 90–92.
Weichhold, G.M., Lautner-Rieske, A. and Zachau, H.G. (1992). *Biol. Chem. Hoppe-Seyler* **373**, 1159–1164.
Weichhold, G.M., Ohnheiser, R. and Zachau, H.G. (1993a). *Genomics* **16**, 503–511.
Weichhold, G.M., Huber, C., Parnes, J.R. and Zachau H.G. (1993b). *Genomics* **16**, 512–514.
Weichhold, G.M., Ohnheiser, R. and Zachau, H.G. (1993c). In 'Progress in Immunology', vol. 8 (J. Gergely *et al.*, eds), pp. 115–119. Springer Verlag, Berlin.
Whitehurst, C., Henney, H.R., Max, E.E., Schroeder, H.W. Jr, Stüber, F., Siminovitch, K.A. and Garrard, W.T. (1992). *Nucleic Acids Res.* **20**, 4929–4930.
Wilson, A.C., Ochman, H. and Prager, E.M. (1987). *Trends Genet.* **3**, 241–247.
Wirth, T., Staudt, L. and Baltimore, D. (1987). *Nature* **329**, 174–178.
Zachau, H.G. (1989a). In 'Immunoglobulin Genes' (T. Honjo, F.W. Alt and T.H. Rabbitts, eds), pp. 91–109. Academic Press, London.
Zachau, H.G. (1989b). In 'Evolutionary Tinkering in Gene Expression' (M. Grunberg-Manago, B.F.C., Clark and H.G. Zachau, eds), pp. 111–119. Plenum Publishing, London.
Zachau, H.G. (1990). *Biol. Chem. Hoppe-Seyler* **371**, 1–6.
Zachau, H.G. (1993). *Gene* **135**, 167–173.
Zachau, H.G. (1995). *The Immunologist*, in press.
Zimmer, F.-J., Huber, C., Quenzel, E.-M., Schek, H., Stiller, C., Thiebe, R. and Zachau, H.G. (1990a). *Biol. Chem. Hoppe-Seyler* **371**, 283–290.
Zimmer, F.-J., Hameister, H., Schek, H. and Zachau, H.G. (1990b). *EMBO J.* **9**, 1535–1542.

Biol. Chem. Hoppe-Seyler
Vol. 374, pp. 1001–1022, November 1993

REVIEW

The Variable Genes of the Human Immunoglobulin κ Locus

Karlheinz F. SCHÄBLE and Hans G. ZACHAU

Institut für Physiologische Chemie der Universität München

(Received 23 August 1993)

Human immunoglobulin genes of the κ type have been studied intensively during the past several years. Since the early work of Bentley and Rabbitts[1,2] a number of laboratories have contributed to the knowledge of the immunoglobulin κ genes and their function. About a dozen years ago, we embarked on a systematic study of the structure of the κ locus, the mechanism of $V_κ$–$J_κ$ rearrangements and the expression of the rearranged genes. The work was reviewed in regular intervals[3–5]. Now that our studies are drawing to a close, a general review was written[6] describing the results and conclusions on the $V_κ$ gene repertoire, on somatic hypermutation, the upstream regulatory elements and the various rearrangements and translocations of the $V_κ$ genes; also some biomedical implications and evolutionary aspects were briefly dealt with. Since probably all $V_κ$ genes of the locus are now known, it seems timely to compile their sequences and to review the conclusions which can be reached from the comparisons.

1) *The structure of the human κ locus*

Fig. 1 gives an overview of the κ locus. It serves as an introduction to this review and, at the same time, presents some results of sequence comparisons which will be discussed below. The locus is largely duplicated comprising a $C_κ$ proximal contig (p) of 600 kb and a distal contig (d) of 440 kb, which are separated by 800 kb of as yet uncloned and probably $V_κ$ gene-free DNA. The detailed restriction map of the locus comprises 3 Mb, while additional 4 Mb were mapped at low resolution[7,8].

The two large contigs p and d emerged from smaller ones that had been studied separately on numerous cosmid and phage λ clones: Op/Od[9], Ap/Ad[10], Lp/Ld[11,12], B[13] and $J_κ$–$C_κ$–κde[14,15]. References to previous reports on the regions are cited in the quoted publications. The small contigs were fused to form the two large ones by chromosomal walking[9–15]. Although we have good indirect evidence that the κ locus has been fully or almost fully cloned[6,12] we try to extend, with the help of YAC clones, the existing contigs at their 5′ and 3′ sides.

2) *Number and classification of the $V_κ$ genes*

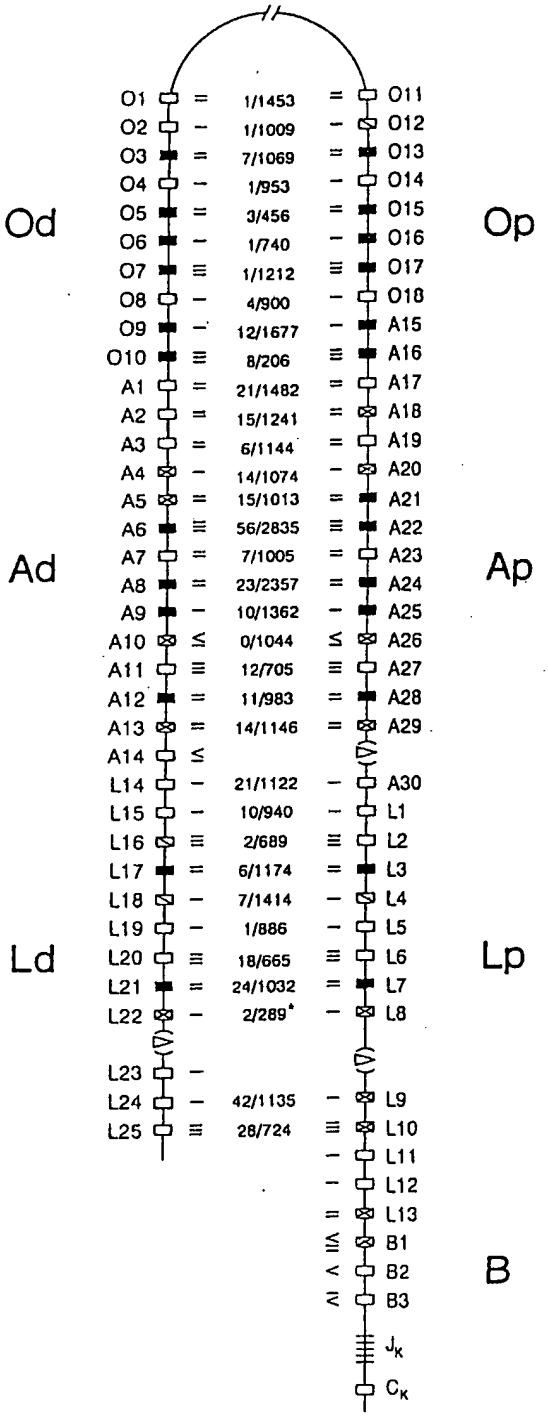
The p contig comprises the single $C_κ$, five $J_κ$ and 40 $V_κ$ gene segments, while in the d contig 36 $V_κ$ gene segments were found. There are 10 solitary $V_κ$ genes and 33 gene pairs whose sequences are 95–100% identical in the coding regions. 32 of the 76 $V_κ$ genes are potentially functional, 16 have minor defects, 25 are pseudogenes and for three genes both potentially functional and slightly defective alleles have been found.

The minor defects are defined as one or two one-bp alterations in a gene, for instance a stop codon in the coding part and/or a deviation from the canonical sequences of regulatory elements, splice sites or the hepta- and nonanucleotide recombination sites. The replacement of a codon for an invariant amino acid (see below) by another codon is not considered to be such a defect, although for some $V_κ$ genes with such alterations no mRNA or κ proteins have been

Abbreviations:
$V_κ$, $J_κ$ and $C_κ$: variable, joining and constant gene segments of the κ locus, respectively; Op/Od, Ap/Ad, Lp/Ld and B regions: $V_κ$ gene-containing regions of the $C_κ$ proximal (p) or distal (d) copies of the human κ locus, respectively (see Fig. 1); bp: base pair(s); kb: kilobase(s) or 1000 bp; Mb: megabase(s) or 10^6 bp; YAC: yeast artificial chromosome; PCR: polymerase chain reaction; pf: potentially functional; mid: minor defects; ps: pseudo; dc: decanucleotide box; pd: 15-mer box.

found[16]. The 16 genes with minor defects are defined as a separate class of genes, since also for them as for the three just mentioned ones potentially functional alleles may exist in the human population.

All known V_κ genes and pseudogenes of the κ locus on chromosome 2p11–12 are listed together with some pertinent data in Table 1. The sequences of all potentially functional and slightly defective V_κ genes and also some pseudogene sequences have been published previously, and the data on the remaining pseudogenes are presented in another report[17]. Also included in Table 1 are the so-called orphon V_κ genes which lie outside the κ locus on the same chromosome or on others. Up to now, 12 orphons were localized on the long arm of chromosome 2, one on its short arm and 11 on other chromosomes. Most of the orphons turned out to be pseudogenes, not only because of their location but also because of defects in their sequences. Two orphon V_κ genes (V108 and V268) have sequences that conform to the criteria of potential functionality; they are, however, listed as pseudogenes in Table 1 because of their location on the long arm of chromosome 2 (V108; ref.[18]) and on the short arm of this chromosome, but on the 3′ side of C_κ at a distance of 1.5 Mb (V268; ref.[19]). There are probably further orphons scattered across the genome[20], but there is no compelling reason for us to study them exhaustively.



◁ Fig. 1. Outline of the human κ locus.

Open boxes represent potentially functional genes, the filled boxes pseudogenes. Open boxes with crossed lines indicate genes with minor defects (as defined in the text). The three genes for which potentially functional *and* slightly defective alleles are known are marked by one diagonal line. The subgroups are indicated by Roman numerals. The deletions (Δ) in the A and L regions have been described[10,11]. The drawing is not to scale. The number of nucleotides sequenced in both the p and d copy gene regions are given together with the number of differences between the regions. An asterisk at L8/L22 indicates that only the part of L22 which is highly homologous to L8[12] is compared.

The evidence that all V_κ genes of the locus have been cloned is indirect. It rests on three arguments.

(i) No V_κ genes were found in the admittedly relatively short extending regions of the existing contigs[9,14].

(ii) Semiquantitative analyses of blot hybridizations of genomic DNA digests with V_κ gene probes indicate that the number of signals not assigned to the cloned V_κ genes at the time is small[21]; by now the genes corresponding to such signals have either been cloned or are attributed to orphon genes[12,20].

(iii) In analyses of the V_κ mRNA/cDNAs and κ proteins all gene products could be assigned to the known germ-line V_κ genes and no evidence for the existence of hitherto undetected germ-line V_κ genes was found[16].

Since the three arguments are circumstantial we can only state that the number of V_κ genes and pseudogenes in the locus very likely is 76.

Table 1. The genes of the human κ locus (011-B3) and the orphon V_κ genes (W1-V268a).

Description[a]	Characteristics[b]	Divergence[c] L	Divergence[c] V	Described in	Accession[d] number
O11	II,St, pf } TP			[9]	X59314
O1	II,St, pf } TP	0	0	[9]	X59311
O11a (V3a)	II,fl, pf	-	3(2)	[9]	X59317
O12	I,St, pf,RP } RTP			[9]	X59315
O2	I,St, pf } RTP	0	0	[9]	X59312
O12a (V3b)	I,fl, md	1(1)	8(3)	[9]	X59318
O13	II,St, ps			[17]	X71888
O3	II,St, ps	-	1	[17]	X71892
O14 (Q5)	I,St, pf } T			[9]	X59316
O4	I,St, pf } T	0	0	[9]	X71893
O4/14a (DILp1)	I,DIL,pf		0	[52]	
O15 (Q8)	II,St, ps			[17]	X71889
O5	II,St, ps	-	0	[17]	X71893
O16 (Q1)	I,St, ps			[17]	X71890
O6	I,St, ps	0	1	[17]	X71894
O6/16a (V55)	I,St, ps	0	2	[53]	X00749
O17	III,St, ps			[17]	X71891
O7	III,St, ps	0	0	[17]	X71895
O18	I,St, pf } RTP			[25]	M64856
O8	I,St, pf } RTP	0	0	[25]	M64855
O18a	I,St, pf	1(1)	0	[25]	M64857
A15	I,St, ps			[17]	X71883
O9	I,St, ps	1	1	[17]	X71896
A15a (HK100)	I,fl, ps	0	0	[1]	V00560
A16	III,St, ps			[17]	X71883
O10	III,St, ps	1	1	[17]	X71896
A16a (Humkv301)	III,fl, ps	-	0	[54]	M17765
A16/O10b (Humkv310)	III,fl, ps	-	0	[54]	M17764
A17	II,St, pf,RTP			[10]	X63403
A1	II,St, pf,T	0	1(1)	[10]	X63402
A18	II,St, md,RT			[10]	X63396
A2	II,St, pf,TP	1	6(5)	[55]	M31952
A19 (Q7)	II, N, pf			[10]	X63397
A3	II,St, pf,RP } RTP	0	0	[56]	X12690
A20 (Y2)	I, N, md,RT			[10]	X63398
A4	I,St, md	2(1)	2(1)	[56]	
A4a (V52)	I,St, md	2(1)	2(1)	[53]	X00748
A21	II,St, md			[17]	X71884
A5	II,St, ps	0	3	[56]	X12689
A22	III,St, ps			[57]	X12685
A6	III,St, ps	1	6	[17]	X71886
A23	II,St, pf,RT			[56]	X12684
A7	II,St, pf	0	4(3)	[10]	X63401
A24	II,St, ps			[17]	X71885
A8	II,St, ps	0	2	[17]	X71887
A25	I,St, ps			[17]	X71885
A9	I,St, ps	1	2	[17]	X71887
A26	VI,St, md } T			[10]	X63399
A10	VI,St, md } T	0	0	[27]	X12683

Description[a]	Characteristics[b]	Divergence[c] L	V	Described in	Accession[d] number
A27	III,St, pf,RTP			[56]	X12686
A27a (Humkv325)	III,fl, pf	0	0	[58]	M15038
A11	III,St, pf,R	0	5(4)	[56]	X12687
A11a (Humkv305)	III,fl, pf,	0	5(4)	~~[55]~~ [54]	M14507
A28	II,St, ps			[56]	X12692
A12	II,St, ps	-	3	[17]	X71882
A29	II,St, md			[10]	X63400
A13	II,St, md	1(1)	8(4)	[10]	X63395
A14	VI,St, md			[27]	X12688
A30	I,fl, pf,RTP			[12]	X72808
L14 (Q4)	I,St, pf,R	0	14(10)	[12]	X63392
L1 (HK137; Q14)	I,fl, pf,TP			[2]	J00248
L15 (HK134; Q13)	I,fl, pf	1	7(4)	[2]	K01323
L15a (HK101)	I,fl, pf	1	8(5)	[1]	V00558
L2 (Humkv328h5;Q11)	III,Hah,p,RT } RTP			[59]	M23090
L16 (Q10)	III,St, p,T }	0	1	[12]	X72815
L16a (Humkv328h2)	III,Hah,pf	0	1	[59]	M23089
L16b (Humkv328)	III,Les,pf	0	1	[59]	M23088
L16c (Humkv329)	III,Les,md	0	3(2)	[59]	M23091
L3 (Q12)	II,Hah,ps			[12]	X72810
L17	II,St, ps	1	3	[12]	X72811
L4 (Va)	I,St, md } P			[60]	X00903
L18 (Va")	I,St, md,T } P	0	1(1)	[61]	X17262
L4/18a (V4a)	I,fl, pf	0	2(2)	[60]	X00900
L5 (Vb)	I,St, pf,T } RTP			[60]	V01577
L19 (Vb")	I,St, pf } RTP	0	1	[61]	X17263
L5/19a (V4b)	I,fl, pf	1	0	[60]	V01576
L6 (Vg)	III,St, pf,RT			[62]	X01668
L20 (Vg")	III,St, pf,R	2(1)	4(3)	[61]	X17264
L6a (38K)	III, pf		0	[63]	
L7 (Vc)	II,St, ps			[60]	X00904
L21 (Vc")	II,St, ps	0	8	[61]	X17265
L8 (Vd)	I,St, md,RT			[60]	X00902
L22	I,St, md	0	-	[12]	X72816
L23 (Q2)	I,GM607,pf			[12]	X72817
L23a	I, N, pf	0	0	[12]	X72818
L9 (Ve)	I,St, md,R			[60]	X00901
L24 (Ve"; Q3; V13)	I,St, pf,RT	1(1)	12(8)	[53]	X00750
L24a	I,N , pf	1(1)	12(8)	[12]	X72819
L10 (Vh)	III,St, md			[62]	X02725
L10a	III, N, md			[12]	X72812
L25 (V138; Q9)	III,St, pf	2(*)	9(5)	[12]	X72820
L11 (Vf)	I,St, pf,T			[25]	M64858
L12 (HK102,V1)	I,fl, pf			[1,53]	V00559
L12a	I,ML, pf,RTP	-	5(1)	[12]	X72813
L13 (Q6)	II, N, md			[12]	X72814

Description[a]	Characteristics[b]	Divergence[c] L	V	Described in	Accession[d] number
B1	VII,St, md			[13]	X12682
B2 (EV15)	V,P1, pf,RT			[64]	X02485
B3	IV,St, pf,RTP			[65]	Z00023
W1	II,St, ps			[66]	X05101
W1a	II,St, ps			[67]	X76074
W2	II,St, ps			[68]	X51884
W3	I,St, ps			[66]	X05102
W4	II,St, ps			[68]	X51883
W5	III,St, ps			[68]	X51882
W6	I,St, ps			[66]	X05103
W7	II,St, ps			[68]	X51881
W8	II,St, ps			[68]	X51880
W9	I,St, ps			[68]	X51879
W10	II,St, ps			[68]	X51886
W11	I,St, ps			[68]	X51885
Chr22-1	I,St, ps			[69]	Z00040
Chr22-2	III,St, ps			[69]	Z00042
Chr22-3	II,St, ps			[69]	Z00041
Chr22-4	II,St, ps			[70]	M20808
Chr22-5	I,St, ps			[53]	X00747
Chr22-5a (V2b)	I,fl, ps			[53]	X00746
Chr1	I,St, ps			[69]	M20809
Z1	I,St, ps			[71]	M23653
Z2	I,St, ps			[20]	X64640
Z3	I,St, ps			[20]	X64641
Z4	I,St, ps			[20]	X64642
V118	I,St, ps			[70]	M20812
V108	I,St, ps(pf)			[18]	X51887
V268	III,St, ps(pf)			[19]	X74459
V268a	III,St, ps(pf)			[19]	X74460

[a] In addition to the current designations previous and alternative ones are given in brackets. The Q/Y nomenclature was used in ref.[21]. Our "standard" alleles isolated from large contigs are named by straight numbers, e.g. L16; a, b, c is added to additional isolates, e.g. L16b. If an allele cannot be assigned to the proximal or distal copy of the κ locus, it receives a double designation, e.g. L5/19a. The C_κ-proximal gene of a pair of duplicates is always listed first.

[b] This column gives several features of the V_κ genes: the subgroups (see text); sources of DNA; fl, fetal liver; St, N, different placenta DNA; GM607, DNA of a lymphoid cell line; the other source designations are taken from the respective references. pf, ps and md refer to potentially functional genes, pseudogenes and genes with minor defects, respectively (see text). The orphon genes V108, V268 and V268a are classified as pseudogenes although they have no obvious defects in their sequence. They are therefore designated ps (pf). R, T and P indicate that for the gene(s) a rearrangement product, a transcription product or a protein is known[16]. The designation is given to a gene pair if the expression product could not be assigned to the p or d copy. For the special situation in the leader of L25 (asterisk) see Fig. 4c and ref.[12].

[c] This column shows the numbers of nucleotides (amino acids in parentheses) in exons 1 and 2 which are divergent between duplicate genes. In the case of alleles the numbers are differences to the p copy genes.

[d] Accession numbers listed are for EMBL and GenBank data libraries.

3) *V_κ gene duplicates and the germ-line repertoire of V_κ genes*

The degree of divergence between duplicate V_κ genes is shown in Table 1 for exons 1 and 2 and in Fig. 1 for the total regions sequenced in both genes of a pair. The average extent of divergence is somewhat higher for pairs of pseudogenes than for pairs of potentially functional genes. In addition, the average divergence

seems to be lower in the O regions than in the A and L regions. Both features may be interpreted in terms of a surveillance mechanism which counteracts the mutational divergence. On the average, the p and d copy-derived sequences of gene regions (Fig. 1) differ by about 1% (404 of 38136bp) indicating that the duplication is an evolutionarily, relatively recent event. 1% divergence is believed to correspond to 10^6 years of evolution[22]. But the differences in the extent of divergence of the various duplicated gene regions will thwart all attempts to date the duplication event exactly.

Because of the generally high similarity among the gene pairs the duplication of the κ locus has not (yet?) increased much the repertoire of different germ-line V_κ genes. Since in most, although not in all cases both genes of a pair belong to the same category being either potentially functional, slightly defective or pseudogenes, the selective advantage of the duplication in allowing the members of a pair to diverge from each other cannot be high. In general, the V_κ genes of the d copy are expressed to a lesser extent than the p copy genes, but there are unequivocal examples of expressed genes of the d copy[7,16,23]. One feature of the duplicated structure may become advantageous in cells that require a second V_κ–J_K rearrangement in order to arrive at a functional joint: since most V_κ genes of the p copy are rearranged by a deletion mechanism the inversional rearrangement of the d copy V_κ genes may be useful in preserving the V_κ gene repertoire[7,23].

The existence of an apparently healthy individual who lacks the d copies of the κ locus in a homozygous fashion demonstrates that the d copy genes are dispensable[7,24]. The absence of one particular d copy gene (A2) in this individual is compensated by V_κ gene(s) of the p copy[25]. This may be true also for other d copy genes since no increase of λ chain-containing antibodies was found in the individual. In serological experiments, kindly performed by *R. Linke*, München, the κ/λ chain ratio was determined to be about 60/40 in both the individual with only the p copies and individuals with p and d copies[26].

4) *Comparison of V_κ genes within subgroups and the derivation of PCR primers*

The alignments of sequences of the subgroups I, II and III are shown in Fig. 2–4. Subgroups IV, V and VII consist of one germ-line V_κ gene each, and subgroup VI has three members, the sequences of which have been compared previously[10,27]. The sequence differences among members of the V_κ gene subgroups and between the subgroups are compiled in Table 2. Matrices of gene-by-gene comparisons are included in ref.[26]. It can be seen in Table 2 that the genes of the V_κIV–V_κVII genes constitute subgroups of their own, albeit single-member subgroups in three cases, and cannot be assigned to the predominant subgroups I–III. It is also confirmed that the V_κ gene subgroups I–IV correspond closely to the classical subgroups I–IV of κ proteins[28] (see below). The differences between the various V_κ gene subgroups are stressed by the comparison of some structural features of the gene regions in Table 3.

The alignments of the V_κ gene sequences of the major subgroups (Fig. 2–4) are an important source of information for all sorts of comparisons and classifications. The alignments also present a rational basis for the selection of PCR primer pairs. A number of primers has been suggested on the basis of the V_κ gene sequences available at the time[29–33]. The arrows in Fig. 2–4 and the comments in the respective legends should be taken as general suggestions. The specific primer sequences to be used depend on the particular experiment. On the 5′ side the choice of primer depends on whether one wants to include the regulatory sequences, the leader region or only exon 2. On the 3′ side one has the choice between the closer or more distant non-coding sequences. In rearranged V_κ–J_κ genes, J_κ consensus or J_κ–C_κ intron sequences are possible, while for cDNAs a part of the C_κ gene may be the sequence of choice. A main determinant in the choice of primers in PCR with genomic DNA is the desired specificity or generality. If one aims at amplifying one specific V_κ gene one may have ro resort to using nested primer pairs (e.g.[12]). Primer pairs that are more or less specific for one of the V_κ gene subgroups can be constructed. Primer pairs or mixtures of different primer pairs for amplifying V_κ genes of all subgroups about equally well are difficult to envisage. The search for the desired primer pairs is facilitated for the reader by the presentation of separate consensus sequences of the potentially functional V_κ genes with and without inclusion of the slightly defective genes.

It was not feasible to include in the alignments the sequences of the 25 pseudogenes of the κ locus and of the 24 known orphon V_κ genes. But depending on the choice of the primer pair one has to expect to amplify in genomic PCR also some of those gene sequences. Another caveat one has to keep in mind for PCR with multigene families as the V_κ genes is the possibility of PCR artifacts. Structures that probably arose as PCR artifacts were described recently for amplification reactions with cDNA mixtures[16], and they conceivably also occur in PCR with genomic DNA when primer pairs of low stringency are employed.

Table 2. Subgroup classification of the V_κ genes of the human κ locus[a].

	I 17,6,6 %	II 9,5,13 %	III 7,1,6 %	IV (B3) 1,-,- %	V (B2) 1,-,- %	VI (A10, A14 A26) -,3,- %	VII (B1) -,1,- %
I							
L	87-100 (>67)						
V	88-100 (>71)						
II							
L	78-92 (>84)	92-100					
V	57-65 (>64)	84-100					
III							
L	68-78	59-67	96-100 (>74)				
V	72-78	66-71 (>68)	92-100				
IV							
L	54-60	55-59	55	100			
V	70-75	68-71	74-78	100			
V							
L	62-68	61-65	63-65	65	100		
V	61-66	57-61	61-65	63	100		
VI							
L	64-79	61-70	61-65	57-60	55-65	77-100	
V	71-73	63-68	70-74	69-70	63	87	
VII							
L	60-64	57-61	53-57	43	59	55-65	100
V	67-70	66-60	68-71	70	65	70-71	100

[a] The numbers underneath the subgroup numerals are the numbers of pf, md and ps genes in the subgroup. L and V refer to exons 1 and 2. The numbers in the Table are percent similarity of potentially functional genes within and between subgroups. The values for slightly defective genes and pseudogenes are given in parentheses.

Table 3. Features of promoter and intron regions of the V_κ gene subgroups.

Subgroup	dc Region[a]		Intron	
	Length [bp]	Similarity[b] [%]	Length[b] [bp]	Similarity[b,c] [%]
I	96–97	81–100 (68–100)	124–126 (108–126)	69–100 (65–100)
II	108–126	83–100 (82–100)	370–440 (333–444)	–
III	103	90–100 (74–100)	170–190 (64–100) ⟷	80–100 (168–1446)
IV	103	–	219	–
V	112	–	145	–
VI	113–115	(80)	206–210	(70)
VII	243	–	107	–

[a] The region from the first nucleotide of dc to exon 1 is compared.
[b] Numbers in parentheses refer to pseudogenes.
[c] Because of the high variability of length no similarities were calculated for the introns of the V_κII-family.

5) Conserved sequence elements

One important outcome of the alignment exercise (Fig. 2–4) is the definition and evaluation of conserved sequence elements. A TATAA-like sequence is found 5′ of all V_κ genes, but its exact sequence is rather variable. The decanucleotide (dc) sequence TNATTTGCAT was recognized early as a functionally important promoter element[34]; independently, the octanucleotide ATTTGCAT was defined as a conserved sequence[35]. The occurrence of the deca- resp. octanucleotide sequence in various gene systems was reviewed[36–38]. In transcription assays the last seven nucleotides of dc appeared to be essential for promoter activity, while alterations in the first and third position of dc allowed reduced transcription[39]. Consequently, deviations from the heptanucleotide TTTGCAT are considered a defect in a V_κ gene. The first three nucleotides of dc seem to be predominantly TGA, TGC or GGA (Fig. 2b, 3b, 4b).

The 15-mer or pd element as defined in ref.[34] is found 17 bp 5′ of dc in all V_κI genes and approximately 150 bp 5′ of dc in A10 and A26, while it is not seen in V_κ genes of other subgroups (Fig. 2a, 3a, 4a). pd is not a promoter element essential for transcription[40] but it may have a supportive activity[41]. Another possibly supportive element is CCCT[41] which is found in one or more copies 20–30 bp 3′ of dc in almost all potentially functional V_κ genes (Fig. 2b, 3b, 4b). An ACCC element, which was found to bind nuclear protein(s)[42], is less well conserved but is found in many potentially functional V_κ genes on the 5′ or 3′ side of dc, in some genes overlapping with the CCCT element. Several sequence motives that have been defined as potential binding sites for transcription factors (review[43]) were found in the sequences by a computer-aided search[44]; they are shown in Fig. 2a and b, 3b, 4b, but appear not to be highly conserved in V_κ genes. In summary, the heptanucleotide part of dc, the TATAA element and the C-rich sequences are highly conserved. pd and other sequences upstream or downstream of dc may help to increase transcriptional activity.

Sequence regularities in the leader and in exon 2 will be discussed on the protein level (see below). Some of the conserved regions in the intron (Fig. 2d, 3d, 4d) are related to functions in the splice process. A remarkable feature are the relatively narrow distribution of intron sizes within the major V_κ gene subgroups and the pronounced differences in this respect between the subgroups (Table 3).

The recombination signal sequences are well conserved in the potentially functional V_κ genes of the major subgroups (Fig. 2f, 3f, 4f). The heptamer agrees fully and the nonamer to a large extent with the canonical sequences. This is so since by our formal definition a deviation from the canonical heptamer sequence is considered a defect that places an otherwise potentially functional gene in the category of genes with minor defects. This does not exclude that such genes do rearrange[16], which is in agreement with the finding that slightly altered recognition sites in a model system lead to a reduced recombination frequency but do not abolish recombination[45]. Considering mechanistic aspects of V_κ–J_κ rearrangements one should keep in mind that in all but very few potentially functional and slightly defective V_κ genes the last codon of exon 2 is a Pro codon; it is always the codon CCT, and even in the few non-Pro codons the last nucleotide is a T. The first nucleotides beyond this codon are, in most genes, C residues. It is an open question whether these nucleotides are recognized by the V_κ–J_κ recombination machinery.

6) Subgroups of κ proteins and the invariant residues in the sequences

It has proven very useful to order immunoglobulins into sets of related sequences which in the case of the human κ chains are called subgroups. Since the early studies of Milstein[46], Hilschmann[47] and others the subgroup classification has become, mainly by the monumentous work of Kabat et al. (ref.[28] and earlier editions) a guiding principle of structural immunology. For subgroups I–IV the classification was fully confirmed when the V_κ gene sequences became known. For five germ-line genes that did not fit into these subgroups the new subgroups V–VII were defined. Transcripts of V_κV and VI genes have been found[32]. Since no proteins are known that are derived from the V_κV–VII genes it may be argued that the one-member families V and VII and the three-member family VI are not subgroups in the sense of the classical definition[28]. But a rigid distinction between sequence families of genes and subgroups of proteins becomes cumbersome, for instance when V_κ gene and protein sequences are compared across species boundaries. An example are the correlations between certain groups of human and murine V_κ genes (exon 2[48]; introns[26]) which are interesting for the consideration of the evolution of the κ locus. It, therefore, appears to be justified to use the subgroup classification also for sequences that are known only at the nucleic acid level.

The formal translation products of the germ-line V_κI–III genes were aligned and are presented in Fig. 5a–c. Sequences derived from potentially functional and slightly defective genes are grouped sepa-

rately. It is not surprising that most residues defined as invariant (i.e. 95% and more conserved within a subgroup) in the collection of protein, cDNA and gene sequence[28] are also found here to be conserved. The relatively high homogeneity of sequences within the subgroups is interesting because of the just mentioned relation between sets of human and mouse genes. Apparently, some subgroup-specific features of V_κ genes have been conserved in evolution for a long time and divergence is counteracted in the species by surveillance mechanisms.

The presentation of the residues conserved among all potentially functional V_κI–IV genes (Fig. 5d) emphasizes the residues that are particularly important for the structure and/or function of the κ chains. Such highly conserved residues are seen all along the sequences and in a cluster at the beginning of FR3. They have to be considered with particular attention in all attempts of antibody gene engineering. The cluster of conserved residues in FR3 may play a role in the intracellular transport and secretion of the protein, as indicated by experiments with mouse λ chains[49]. Its importance is also seen from the fact that it is not only conserved between the κ subgroups and between κ and λ chains but also in several mammalian species. The same is true for the four residues shown with double underlining in Fig. 5d. Their conservation is illustrated by the numbers of light chains in which they do or do not occur: Cys 23, 1450/3; Trp 35, 1242/3; Gly 57, 1133/12; Cys 88, 1134/1. The numbers are derived from the sequence collection of Kabat et al.[28]

The leader sequence that functions in intracellular transport and/or secretion of the proteins is highly conserved within the subgroups, while between subgroups only some hydrophobic residues in the 3′ part of the leader sequence are kept invariant.

7) Some aspects of the evolution of the human κ locus

To discuss evolution is one of the attractions of studying a multigene familiy like the V_κ genes. Such discussions play a role in many reports from other laboratories (e.g.[2,28,33]) and from our laboratory. Some of our earlier findings were discussed in the context of evolution in a review article[50], for instance the duplication of the locus, the interdigitation of genes of different subgroups and the relationship between orphons and V_κ genes of the locus. While the evolution of most pseudogenes and orphons is dwelled on in another report[17] we like to enumerate here some of the reactions or mechanisms that must have operated in shaping the present-day κ locus; we will then turn to the discussion of sequence similarities between V_κ genes and groups of such genes.

After duplication, interdigitation, conservation of invariant residues and generation of orphons the point mutations should be mentioned next. The average divergence of about 1% between the p and d copies which has been mentioned in section 3 is largely due to point mutations. Not surprisingly, there are about twice as many transitions than transversions, and no indication for a pronounced selection was detected[26]. Stop codons have been created in the locus 18 times, some of them in genes which are pseudogenes also on other accounts. The insertion of a single G residue in exon 2 of the gene A4 led to a debilitating frame shift[10]. Clear examples of gene conversion were observed between some genes of the L region[2,12] and indications of information transfer are also seen in other regions of the locus. Deletions of gene-containing stretches of DNA could well be defined in cases where they occurred only in one of the copies of the locus[10], or in both copies but with different breakpoints of homology[11]. Insertions of repetitive elements occurred both before and, as in the case of one particular Alu element[10], after the duplication.

The similarities between the V_κ genes of subgroup I are particularly high in the L regions, in part certainly because of the just mentioned gene conversions. The homologies extend far into the downstream region to a particular structural element, the so-called L element[51], which is found only in this group of V_κI genes and in the orphon V_κI gene V108[18]. The V_κI genes of the O regions seem to form, together with some orphon V_κI genes, a group of their own. The V_κ genes of subgroup II are less similar to each other than the ones of other subgroups. Also here the L region and the O region plus some orphon V_κII genes form groups of their own. The V_κIII genes again constitute a somewhat more homogeneous group with the possibility to define more closely related subsets in the L, A, and O regions. The relationships between the genes of the three main subgroups are discussed in some detail in ref.[26]. Relationship diagrams, which can be drawn in all cases, have to be taken with a grain or even a kilogram of salt if it comes to equating them with evolutionary trees, not the least because of gene conversion and possible other surveillance mechanisms. The same is true for the ur-V_κ gene and the ur-V_κI–IV (or I–VII) V_κ genes which one somehow has to postulate. The comparison of human and mouse V_κ gene sequences indicates that, at least in some cases, subgroup formation has predated speciation[26,48]. The evolution of the human κ locus will become better understandable once the κ loci of primate(s) and other mammals are known.

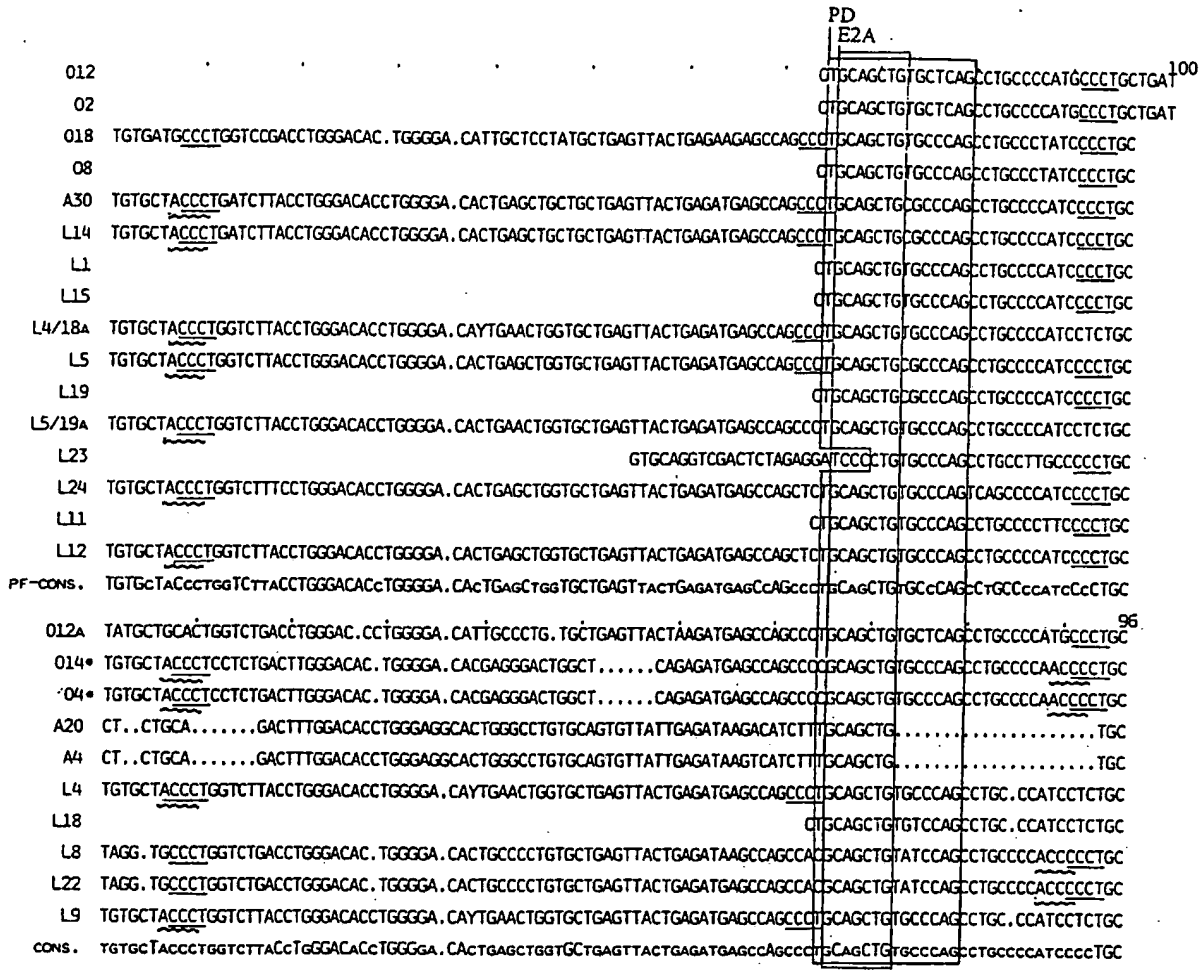


Fig. 2a

Fig. 2. Alignment of the gene regions of the potentially functional (pf) and slightly defective (md) V_x genes of subgroup I.

For the boxed and underlined elements see text. Alleles of genes are shown only if they differ from their respective standard gene. The upper blocks of sequences show all pf genes and the lower one the genes with minor defects. Consensus sequences are given for the pf genes (pf-cons.) and for pf plus md genes together (cons.). Lower case letters are used for nucleotides occuring in more than 50% of the positions. A dash indicates that no nucleotide occurs in more than 50% of the sequences at the position. Dotted lines indicate gaps found in the sequences. Gene sequences having no obvious defects at the nucleotide level, but showing exchanges of invariant amino acids are placed with the md genes. They are marked by an asterisk.

a) Approximately 100 bp upstream of the dc box are shown. The pd box would be suitable as part of a largely V_xI-specific primer.

b) The 100 bp upstream of the leader sequence are shown. For the boxed and underlined elements see text. L9 is considered to be a gene with "minor" defects because of a deletion in the dc box and an altered first ATG in exon 1. The arrow designates one of the possible V_xI-specific primers.

c) Exon 1. The 5′ part of the arrow may yield a largely V_xI- plus V_xII-specific primer while the 3′ part seems to be more or less V_xI- plus V_xIII-specific.

d) Intron sequences. The possible primer may have a limited subgroup specificity.

e) Exon 2. In the framework regions several primers are conceivable.

f) Where known, the first 100 bp downstream of exon 2 are shown. The hepta- and nonanucleotide recognition sites are underlined. It may be difficult, because of these sites, to construct subgroup-specific primers.

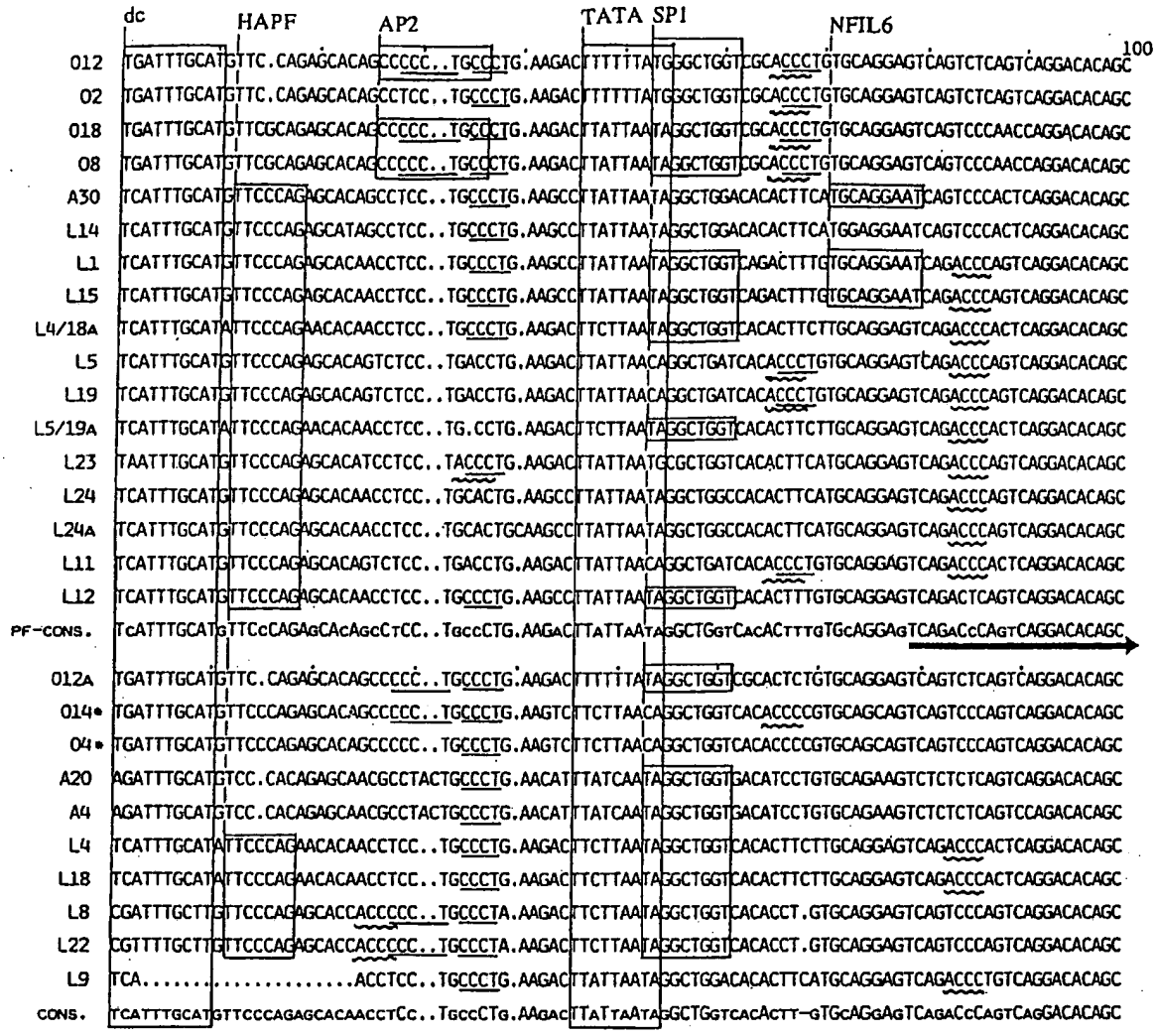

dc
HAPF
AP2
TATA
SP1
NFIL6
100
012 TGATTTGCATGTTC.CAGAGCACAGCCCCC..TGCCCTG.AAGACTTTTTTATGGGCTGGTCGCACCCTGTGCAGGAGTCAGTCTCAGTCAGGACACAGC
02 TGATTTGCATGTTC.CAGAGCACAGCCTCC..TGCCCTG.AAGACTTTTTTATGGGCTGGTCGCACCCTGTGCAGGAGTCAGTCTCAGTCAGGACACAGC
018 TGATTTGCATGTTCGCAGAGCACAGCCCCC..TGCCCTG.AAGACTTATTAATAGGCTGGTCGCACCCTGTGCAGGAGTCAGTCCCAACCAGGACACAGC
08 TGATTTGCATGTTCGCAGAGCACAGCCCCC..TGCCCTG.AAGACTTATTAATAGGCTGGTCGCACCCTGTGCAGGAGTCAGTCCCAACCAGGACACAGC
A30 TCATTTGCATGTTCCCAGAGCACAGCCTCC..TGCCCTG.AAGCCTTATTAATAGGCTGGACACACTTCATGCAGGAATCAGTCCCACTCAGGACACAGC
L14 TCATTTGCATGTTCCCAGAGCATAGCCTCC..TGCCCTG.AAGCCTTATTAATAGGCTGGACACACTTCATGGAGGAATCAGTCCCACTCAGGACACAGC
L1 TCATTTGCATGTTCCCAGAGCACAACCTCC..TGCCCTG.AAGCCTTATTAATAGGCTGGTCAGACTTTGTGCAGGAATCAGACCCAGTCAGGACACAGC
L15 TCATTTGCATGTTCCCAGAGCACAACCTCC..TGCCCTG.AAGCCTTATTAATAGGCTGGTCAGACTTTGTGCAGGAATCAGACCCAGTCAGGACACAGC
L4/18A TCATTTGCATATTCCCAGAACACAACCTCC..TGCCCTG.AAGACTTCTTAATAGGCTGGTCACACTTCTTGCAGGAGTCAGACCCACTCAGGACACAGC
L5 TCATTTGCATGTTCCCAGAGCACAGTCTCC..TGACCTG.AAGACTTATTAACAGGCTGATCACACCCTGTGCAGGAGTCAGACCCAGTCAGGACACAGC
L19 TCATTTGCATGTTCCCAGAGCACAGTCTCC..TGACCTG.AAGACTTATTAACAGGCTGATCACACCCTGTGCAGGAGTCAGACCCAGTCAGGACACAGC
L5/19A TCATTTGCATATTCCCAGAACACAACCTCC..TG.CCTG.AAGACTTCTTAATAGGCTGGTCACACTTCTTGCAGGAGTCAGACCCACTCAGGACACAGC
L23 TAATTTGCATGTTCCCAGAGCACATCCTCC..TACCCTG.AAGACTTATTAATGCGCTGGTCACACTTCATGCAGGAGTCAGACCCAGTCAGGACACAGC
L24 TCATTTGCATGTTCCCAGAGCACAACCTCC..TGCACTG.AAGCCTTATTAATAGGCTGGCCACACTTCATGCAGGAGTCAGACCCAGTCAGGACACAGC
L24A TCATTTGCATGTTCCCAGAGCACAACCTCC..TGCACTGCAAGCCTTATTAATAGGCTGGCCACACTTCATGCAGGAGTCAGACCCAGTCAGGACACAGC
L11 TCATTTGCATGTTCCCAGAGCACAGTCTCC..TGACCTG.AAGACTTATTAACAGGCTGATCACACCCTGTGCAGGAGTCAGACCCACTCAGGACACAGC
L12 TCATTTGCATGTTCCCAGAGCACAACCTCC..TGCCCTG.AAGCCTTATTAATAGGCTGGTCACACTTTGTGCAGGAGTCAGACTCAGTCAGGACACAGC
PF-CONS. TCATTTGCATGTTCCCAGAGCACAGCCTCC..TGCCCTG.AAGACTTATTAATAGGCTGGTCACACTTTGTGCAGGAGTCAGACCCAGTCAGGACACAGC
012A TGATTTGCATGTTC.CAGAGCACAGCCCCC..TGCCCTG.AAGACTTTTTTATAGGCTGGTCGCACTCTGTGCAGGAGTCAGTCTCAGTCAGGACACAGC
014* TGATTTGCATGTTCCCAGAGCACAGCCCCC..TGCCCTG.AAGTCTTCTTAACAGGCTGGTCACACCCCGTGCAGCAGTCAGTCCCAGTCAGGACACAGC
04* TGATTTGCATGTTCCCAGAGCACAGCCCCC..TGCCCTG.AAGTCTTCTTAACAGGCTGGTCACACCCCGTGCAGCAGTCAGTCCCAGTCAGGACACAGC
A20 AGATTTGCATGTCC.CACAGAGCAACGCCTACTGCCCTG.AACATTTATCAATAGGCTGGTGACATCCTGTGCAGAAGTCTCTCTCAGTCAGGACACAGC
A4 AGATTTGCATGTCC.CACAGAGCAACGCCTACTGCCCTG.AACATTTATCAATAGGCTGGTGACATCCTGTGCAGAAGTCTCTCTCAGTCCAGACACAGC
L4 TCATTTGCATATTCCCAGAACACAACCTCC..TGCCCTG.AAGACTTCTTAATAGGCTGGTCACACTTCTTGCAGGAGTCAGACCCACTCAGGACACAGC
L18 TCATTTGCATATTCCCAGAACACAACCTCC..TGCCCTG.AAGACTTCTTAATAGGCTGGTCACACTTCTTGCAGGAGTCAGACCCACTCAGGACACAGC
L8 CGATTTGCTTGTTCCCAGAGCACCACCCCC..TGCCCTA.AAGACTTCTTAATAGGCTGGTCACACCT.GTGCAGGAGTCAGTCCCAGTCAGGACACAGC
L22 CGTTTTGCTTGTTCCCAGAGCACCACCCCC..TGCCCTA.AAGACTTCTTAATAGGCTGGTCACACCT.GTGCAGGAGTCAGTCCCAGTCAGGACACAGC
L9 TCA.....ACCTCC..TGCCCTG.AAGACTTATTAATAGGCTGGACACACTTCATGCAGGAGTCAGACCCTGTCAGGACACAGC
CONS. TCATTTGCATGTTCCCAGAGCACAACCTCC..TGCCCTG.AAGACTTATTAATAGGCTGGTCACACTT-GTGCAGGAGTCAGACCCAGTCAGGACACAGC

Fig. 2b

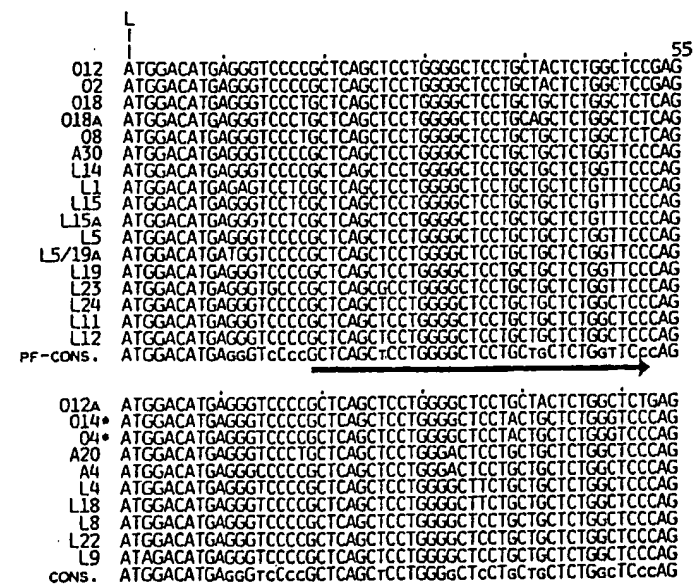

L
55
012 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTACTCTGGCTCCGAG
02 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTACTCTGGCTCCGAG
018 ATGGACATGAGGGTCCCTGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCTCAG
018A ATGGACATGAGGGTCCCTGCTCAGCTCCTGGGGCTCCTGCAGCTCTGGCTCTCAG
08 ATGGACATGAGGGTCCCTGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCTCAG
A30 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGTTCCCAG
L14 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGTTCCCAG
L1 ATGGACATGAGAGTCCTCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGTTTCCCAG
L15 ATGGACATGAGGGTCCTCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGTTTCCCAG
L15A ATGGACATGAGGGTCCTCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGTTTCCCAG
L5 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGTTCCCAG
L5/19A ATGGACATGATGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGTTCCCAG
L19 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGTTCCCAG
L23 ATGGACATGAGGGTGCCCGCTCAGCGCCTGGGGCTCCTGCTGCTCTGGTTCCCAG
L24 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCCCAG
L11 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCCCAG
L12 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCCCAG
PF-CONS. ATGGACATGAGGGTcCccGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGTTCccAG
012A ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTACTCTGGCTCTGAG
014• ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTACTGCTCTGGGTCCCAG
04• ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTACTGCTCTGGGTCCCAG
A20 ATGGACATGAGGGTCCCTGCTCAGCTCCTGGGACTCCTGCTGCTCTGGCTCCCAG
A4 ATGGACATGAGGGCCCCCGCTCAGCTCCTGGGACTCCTGCTGCTCTGGCTCCCAG
L4 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTTCTGCTGCTCTGGCTCCCAG
L18 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTTCTGCTGCTCTGGCTCCCAG
L8 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCCCAG
L22 ATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCCCAG
L9 ATAGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCCCAG
CONS. ATGGACATGAGGGTcCccGCTCAGCTCCTGGGGCTcCTGCTGCTCTGGcTCccAG

Fig. 2c

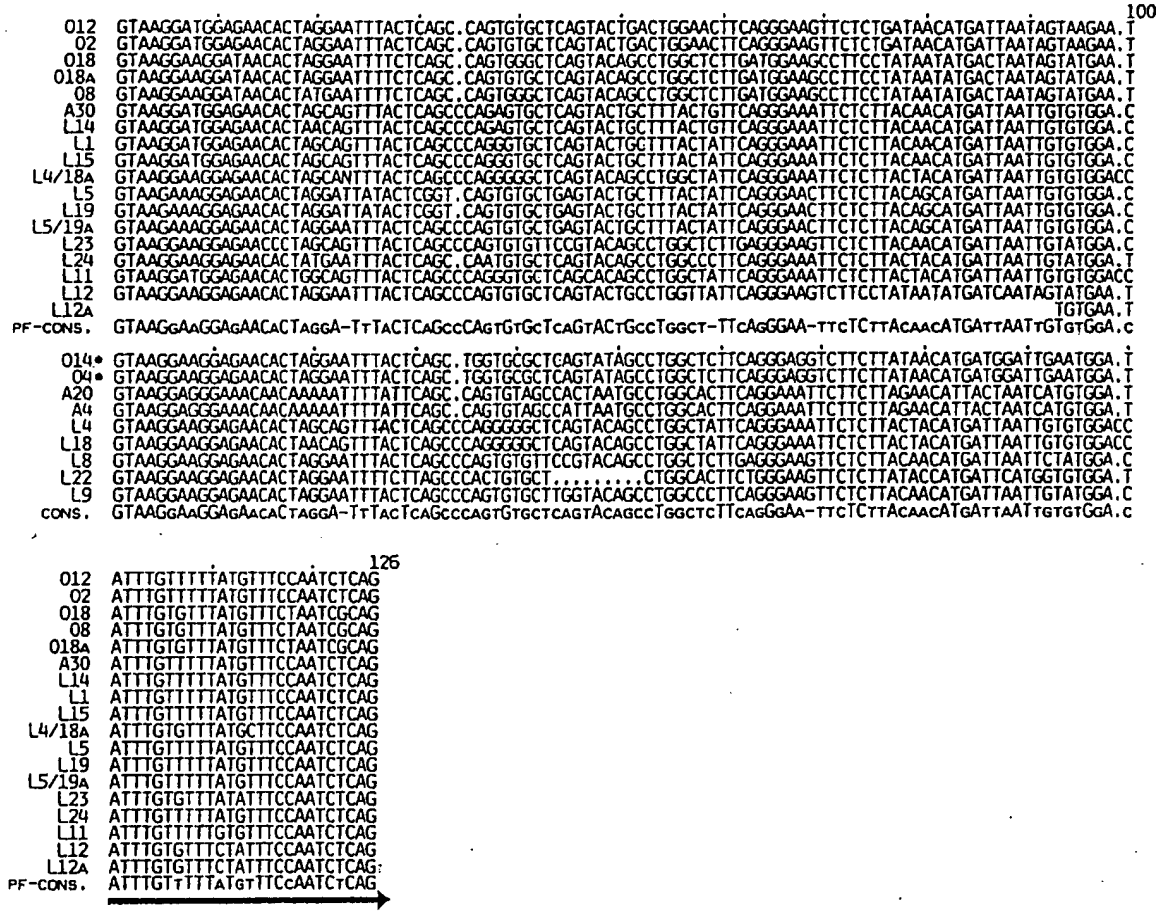

100
012 GTAAGGATGGAGAACACTAGGAATTTACTCAGC.CAGTGTGCTCAGTACTGACTGGAACTTCAGGGAAGTTCTCTGATAACATGATTAATAGTAAGAA.T
02 GTAAGGATGGAGAACACTAGGAATTTACTCAGC.CAGTGTGCTCAGTACTGACTGGAACTTCAGGGAAGTTCTCTGATAACATGATTAATAGTAAGAA.T
018 GTAAGGAAGGATAACACTAGGAATTTTCTCAGC.CAGTGTGCTCAGTACAGCCTGGCTCTTGATGGAAGCCTTCCTATAATATGACTAATAGTATGAA.T
018A GTAAGGAAGGATAACACTAGGAATTTTCTCAGC.CAGTGTGCTCAGTACAGCCTGGCTCTTGATGGAAGCCTTCCTATAATATGACTAATAGTATGAA.T
08 GTAAGGAAGGATAACACTATGAATTTTCTCAGC.CAGTGGGCTCAGTACAGCCTGGCTCTTGATGGAAGCCTTCCTATAATATGACTAATAGTATGAA.T
A30 GTAAGGATGGAGAACACTAGCAGTTTACTCAGCCCAGAGTGCTCAGTACTGCTTTACTGTTCAGGGAAATTCTCTTACAACATGATTAATTGTGTGGA.C
L14 GTAAGGATGGAGAACACTAACAGTTTACTCAGCCCAGAGTGCTCAGTACTGCTTTACTGTTCAGGGAAATTCTCTTACAACATGATTAATTGTGTGGA.C
L1 GTAAGGATGGAGAACACTAGCAGTTTACTCAGCCCAGGGTGCTCAGTACTGCTTTACTATTCAGGGAAATTCTCTTACAACATGATTAATTGTGTGGA.C
L15 GTAAGGATGGAGAACACTAGCAGTTTACTCAGCCCAGGGTGCTCAGTACTGCTTTACTATTCAGGGAAATTCTCTTACAACATGATTAATTGTGTGGA.C
L4/18A GTAAGGAAGGAGAACACTAGCANTTTACTCAGCCCAGGGGGCTCAGTACAGCCTGGCTATTCAGGGAAATTCTCTTACTACATGATTAATTGTGTGGACC
L5 GTAAGAAAGGAGAACACTAGGATTATACTCGGT.CAGTGTGCTGAGTACTGCTTTACTATTCAGGGAACTTCTCTTACAGCATGATTAATTGTGTGGA.C
L19 GTAAGAAAGGAGAACACTAGGATTATACTCGGT.CAGTGTGCTGAGTACTGCTTTACTATTCAGGGAACTTCTCTTACAGCATGATTAATTGTGTGGA.C
L5/19A GTAAGAAAGGAGAACACTAGGAATTTACTCAGCCCAGTGTGCTGAGTACTGCTTTACTATTCAGGGAACTTCTCTTACAGCATGATTAATTGTGTGGA.C
L23 GTAAGGAAGGAGAACCCTAGCAGTTTACTCAGCCCAGTGTGTTCCGTACAGCCTGGCTCTTGAGGGAAGTTCTCTTACAACATGATTAATTGTATGGA.C
L24 GTAAGGAAGGAGAACACTATGAATTTACTCAGC.CAATGTGCTCAGTACAGCCTGGCCCTTCAGGGAAATTCTCTTACTACATGATTAATTGTATGGA.T
L11 GTAAGGATGGAGAACACTGGCAGTTTACTCAGCCCAGGGTGCTCAGCACAGCCTGGCTATTCAGGGAAATTCTCTTACTACATGATTAATTGTGTGGACC
L12 GTAAGGAAGGAGAACACTAGGAATTTACTCAGCCCAGTGTGCTCAGTACTGCCTGGTTATTCAGGGAAGTCTTCCTATAATATGATCAATAGTATGAA.T
L12A TGTGAA.T
PF-CONS. GTAAGgAaGGAgAACaCTAggA-TtTaCTCaGccCAgtGtGcTcaGtACtGccTggct-TTcAgGGAA-ttcTCttAcaacATGAttAATtGTgtGgA.c
014• GTAAGGAAGGAGAACACTAGGAATTTACTCAGC.TGGTGCGCTCAGTATAGCCTGGCTCTTCAGGGAGGTCTTCTTATAACATGATGGATTGAATGGA.T
04• GTAAGGAAGGAGAACACTAGGAATTTACTCAGC.TGGTGCGCTCAGTATAGCCTGGCTCTTCAGGGAGGTCTTCTTATAACATGATGGATTGAATGGA.T
A20 GTAAGGAGGGAAACAACAAAAATTTTATTCAGC.CAGTGTAGCCACTAATGCCTGGCACTTCAGGAAATTCTTCTTAGAACATTACTAATCATGTGGA.T
A4 GTAAGGAGGGAAACAACAAAAATTTTATTCAGC.CAGTGTAGCCATTAATGCCTGGCACTTCAGGAAATTCTTCTTAGAACATTACTAATCATGTGGA.T
L4 GTAAGGAAGGAGAACACTAGCAGTTTACTCAGCCCAGGGGGCTCAGTACAGCCTGGCTATTCAGGGAAATTCTCTTACTACATGATTAATTGTGTGGACC
L18 GTAAGGAAGGAGAACACTAACAGTTTACTCAGCCCAGGGGGCTCAGTACAGCCTGGCTATTCAGGGAAATTCTCTTACTACATGATTAATTGTGTGGACC
L8 GTAAGGAAGGAGAACACTAGCAGTTTACTCAGCCCAGTGTGTTCCGTACAGCCTGGCTCTTGAGGGAAGTTCTCTTACAACATGATTAATTCTATGGA.C
L22 GTAAGGAAGGAGAACACTAGGAATTTTCTTAGCCCACTGTGCT.....CTGGCACTTCTGGGAAGTTCTCTTATACCATGATTCATGGTGTGGA.T
L9 GTAAGGAAGGAGAACACTAGGAATTTACTCAGCCCAGTGTGCTTGGTACAGCCTGGCCCTTCAGGGAAGTTCTCTTACAACATGATTAATTGTATGGA.C
CONS. GTAAGgAaGGAgAacaCTaggA-TtTaCTcaGcccagtGtgctcagtAcagccTggctcTTcaggAA-ttcTCttAcaacATgAttAATtgtgtGgA.c
126
012 ATTTGTTTTTATGTTTCCAATCTCAG
02 ATTTGTTTTTATGTTTCCAATCTCAG
018 ATTTGTGTTTATGTTTCTAATCGCAG
08 ATTTGTGTTTATGTTTCTAATCGCAG
018A ATTTGTGTTTATGTTTCTAATCGCAG
A30 ATTTGTTTTTATGTTTCCAATCTCAG
L14 ATTTGTTTTTATGTTTCCAATCTCAG
L1 ATTTGTTTTTATGTTTCCAATCTCAG
L15 ATTTGTTTTTATGTTTCCAATCTCAG
L4/18A ATTTGTGTTTATGCTTCCAATCTCAG
L5 ATTTGTTTTTATGTTTCCAATCTCAG
L19 ATTTGTTTTTATGTTTCCAATCTCAG
L5/19A ATTTGTTTTTATGTTTCCAATCTCAG
L23 ATTTGTGTTTATATTTCCAATCTCAG
L24 ATTTGTTTTTATGTTTCCAATCTCAG
L11 ATTTGTTTTTGTGTTTCCAATCTCAG
L12 ATTTGTGTTTCTATTTCCAATCTCAG
L12A ATTTGTGTTTCTATTTCCAATCTCAG
PF-CONS. ATTTGTtTTTAtGtTTCcAATCtCAG
014• GTTTGTTTTTTATATTTCCAATCTCAG
04• GTTTGTTTTTTATATTTCCAATCTCAG
A20 ATGTGTTTTTTATGTTCCTAATATCAG
A4 ATGTGTTTTTTATGTTCCTAATATCAG
L4 ATTTGTGTTTATGCTTCCAATCTCAG
L18 ATTTGTGTTTATGCTTCCAATCTCAG
L8 ATTTGTGTTTATATTTCCAATCTCAG
L22 ATTTGTTTTTTATGTTTCCAATCTCAG
L9 ATTTGTTTTTTATGTTTCCAATCTCAG
CONSERV. ATtTGTtTTTAtGtTTCcAATCtCAG

Fig. 2d

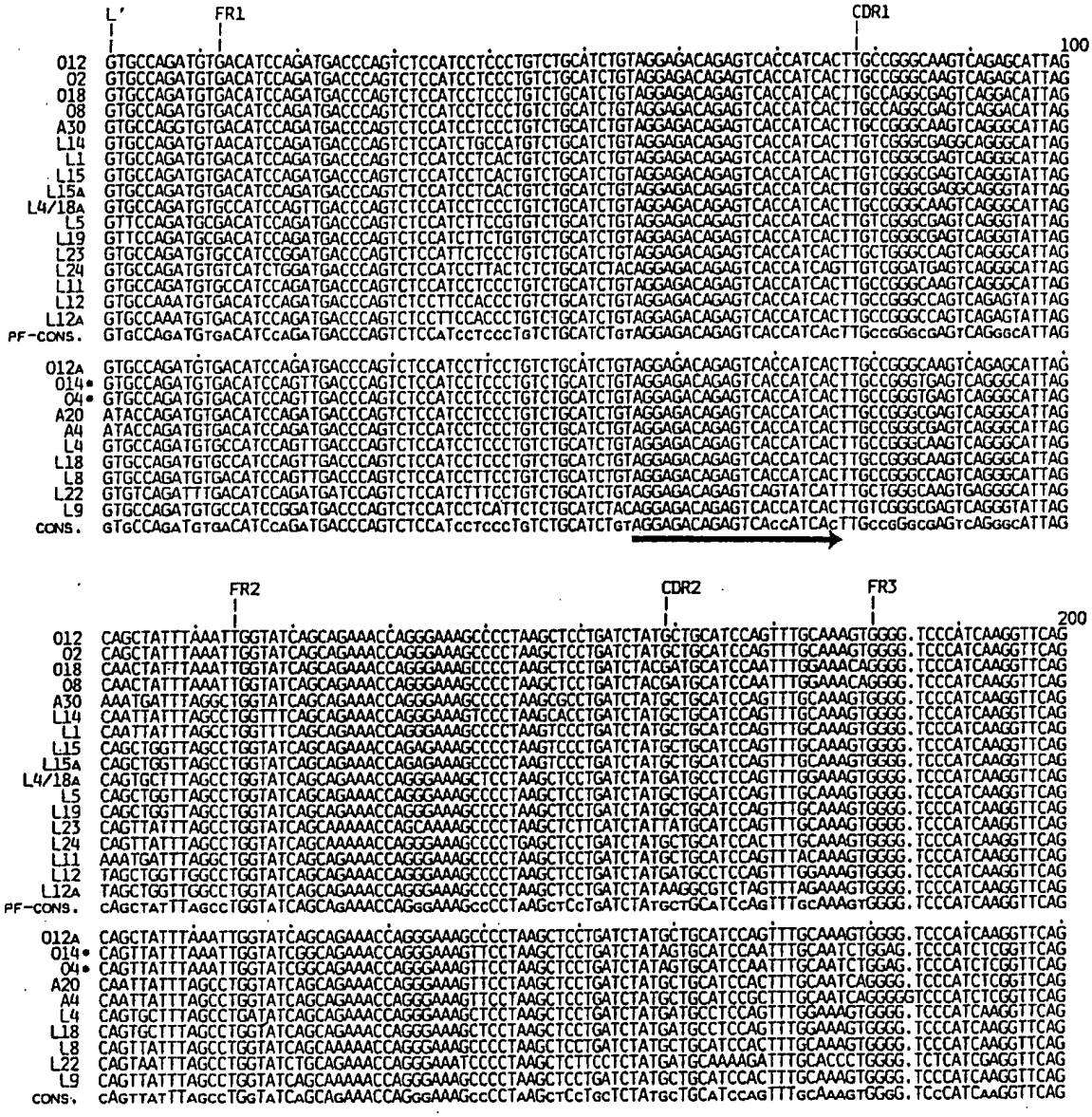

L'
FR1
CDR1
100
012 GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCAAGTCAGAGCATTAG
02 GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCAAGTCAGAGCATTAG
018 GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCAGGCGAGTCAGGACATTAG
08 GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCAGGCGAGTCAGGACATTAG
A30 GTGCCAGGTGTGACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCAAGTCAGGGCATTAG
L14 GTGCCAGATGTAACATCCAGATGACCCAGTCTCCATCTGCCATGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGTCGGGCGAGGCAGGGCATTAG
L1 GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCACTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGTCGGGCGAGTCAGGGCATTAG
L15 GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCACTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGTCGGGCGAGTCAGGGTATTAG
L15A GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCACTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGTCGGGCGAGGCAGGGTATTAG
L4/18A GTGCCAGATGTGCCATCCAGTTGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCAAGTCAGGGCATTAG
L5 GTTCCAGATGCGACATCCAGATGACCCAGTCTCCATCTTCCGTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGTCGGGCGAGTCAGGGTATTAG
L19 GTTCCAGATGCGACATCCAGATGACCCAGTCTCCATCTTCTGTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGTCGGGCGAGTCAGGGTATTAG
L23 GTGCCAGATGTGCCATCCGGATGACCCAGTCTCCATTCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCTGGGCCAGTCAGGGCATTAG
L24 GTGCCAGATGTGTCATCTGGATGACCCAGTCTCCATCCTTACTCTCTGCATCTACAGGAGACAGAGTCACCATCAGTTGTCGGATGAGTCAGGGCATTAG
L11 GTGCCAGATGTGCCATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCAAGTCAGGGCATTAG
L12 GTGCCAAATGTGACATCCAGATGACCCAGTCTCCTTCCACCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCCAGTCAGAGTATTAG
L12A GTGCCAAATGTGACATCCAGATGACCCAGTCTCCTTCCACCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCCAGTCAGAGTATTAG
PF-CONS. GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCGAGTCAGGGCATTAG
012A GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCTTCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCAAGTCAGAGCATTAG
014• GTGCCAGATGTGACATCCAGTTGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGTGAGTCAGGGCATTAG
04• GTGCCAGATGTGACATCCAGTTGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGTGAGTCAGGGCATTAG
A20 ATACCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCGAGTCAGGGCATTAG
A4 ATACCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCGAGTCAGGGCATTAG
L4 GTGCCAGATGTGCCATCCAGTTGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCAAGTCAGGGCATTAG
L18 GTGCCAGATGTGCCATCCAGTTGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCAAGTCAGGGCATTAG
L8 GTGCCAGATGTGACATCCAGTTGACCCAGTCTCCATCCTTCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCCAGTCAGGGCATTAG
L22 GTGTCAGATTTGACATCCAGATGATCCAGTCTCCATCTTTCCTGTCTGCATCTGTAGGAGACAGAGTCAGTATCATTTGCTGGGCAAGTGAGGGCATTAG
L9 GTGCCAGATGTGCCATCCGGATGACCCAGTCTCCATCCTCATTCTCTGCATCTACAGGAGACAGAGTCACCATCACTTGTCGGGCGAGTCAGGGTATTAG
CONS. GTGCCAGATGTGACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGACAGAGTCACCATCACTTGCCGGGCGAGTCAGGGCATTAG
FR2
CDR2
FR3
200
012 CAGCTATTTAAATTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
02 CAGCTATTTAAATTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
018 CAACTATTTAAATTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTACGATGCATCCAATTTGGAAACAGGGG.TCCCATCAAGGTTCAG
08 CAACTATTTAAATTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTACGATGCATCCAATTTGGAAACAGGGG.TCCCATCAAGGTTCAG
A30 AAATGATTTAGGCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCGCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L14 CAATTATTTAGCCTGGTTTCAGCAGAAACCAGGGAAAGTCCCTAAGCACCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L1 CAATTATTTAGCCTGGTTTCAGCAGAAACCAGGGAAAGCCCCTAAGTCCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L15 CAGCTGGTTAGCCTGGTATCAGCAGAAACCAGAGAAAGCCCCTAAGTCCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L15A CAGCTGGTTAGCCTGGTATCAGCAGAAACCAGAGAAAGCCCCTAAGTCCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L4/18A CAGTGCTTTAGCCTGGTATCAGCAGAAACCAGGGAAAGCTCCTAAGCTCCTGATCTATGATGCCTCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L5 CAGCTGGTTAGCCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L19 CAGCTGGTTAGCCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L23 CAGTTATTTAGCCTGGTATCAGCAAAAACCAGCAAAAGCCCCTAAGCTCTTCATCTATTATGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L24 CAGTTATTTAGCCTGGTATCAGCAAAAACCAGGGAAAGCCCCTGAGCTCCTGATCTATGCTGCATCCACTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L11 AAATGATTTAGGCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTACAAAGTGGGG.TCCCATCAAGGTTCAG
L12 TAGCTGGTTGGCCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGATGCCTCCAGTTTGGAAAGTGGGG.TCCCATCAAGGTTCAG
L12A TAGCTGGTTGGCCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATAAGGCGTCTAGTTTAGAAAGTGGGG.TCCCATCAAGGTTCAG
PF-CONS. CAGCTATTTAGCCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
012A CAGCTATTTAAATTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
014• CAGTTATTTAAATTGGTATCGGCAGAAACCAGGGAAAGTTCCTAAGCTCCTGATCTATAGTGCATCCAATTTGCAATCTGGAG.TCCCATCTCGGTTCAG
04• CAGTTATTTAAATTGGTATCGGCAGAAACCAGGGAAAGTTCCTAAGCTCCTGATCTATAGTGCATCCAATTTGCAATCTGGAG.TCCCATCTCGGTTCAG
A20 CAATTATTTAGCCTGGTATCAGCAGAAACCAGGGAAAGTTCCTAAGCTCCTGATCTATGCTGCATCCACTTTGCAATCAGGGG.TCCCATCTCGGTTCAG
A4 CAATTATTTAGCCTGGTATCAGCAGAAACCAGGGAAAGTTCCTAAGCTCCTGATCTATGCTGCATCCGCTTTGCAATCAGGGGGTCCCATCTCGGTTCAG
L4 CAGTGCTTTAGCCTGATATCAGCAGAAACCAGGGAAAGCTCCTAAGCTCCTGATCTATGATGCCTCCAGTTTGGAAAGTGGGG.TCCCATCAAGGTTCAG
L18 CAGTGCTTTAGCCTGGTATCAGCAGAAACCAGGGAAAGCTCCTAAGCTCCTGATCTATGATGCCTCCAGTTTGGAAAGTGGGG.TCCCATCAAGGTTCAG
L8 CAGTTATTTAGCCTGGTATCAGCAAAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCACTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
L22 CAGTAATTTAGCCTGGTATCTGCAGAAACCAGGGAAATCCCCTAAGCTCTTCCTCTATGATGCAAAAGATTTGCACCCTGGGG.TCTCATCGAGGTTCAG
L9 CAGTTATTTAGCCTGGTATCAGCAAAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCACTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG
CONS. CAGTTATTTAGCCTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGCTCTATGCTGCATCCAGTTTGCAAAGTGGGG.TCCCATCAAGGTTCAG

Fig. 2e

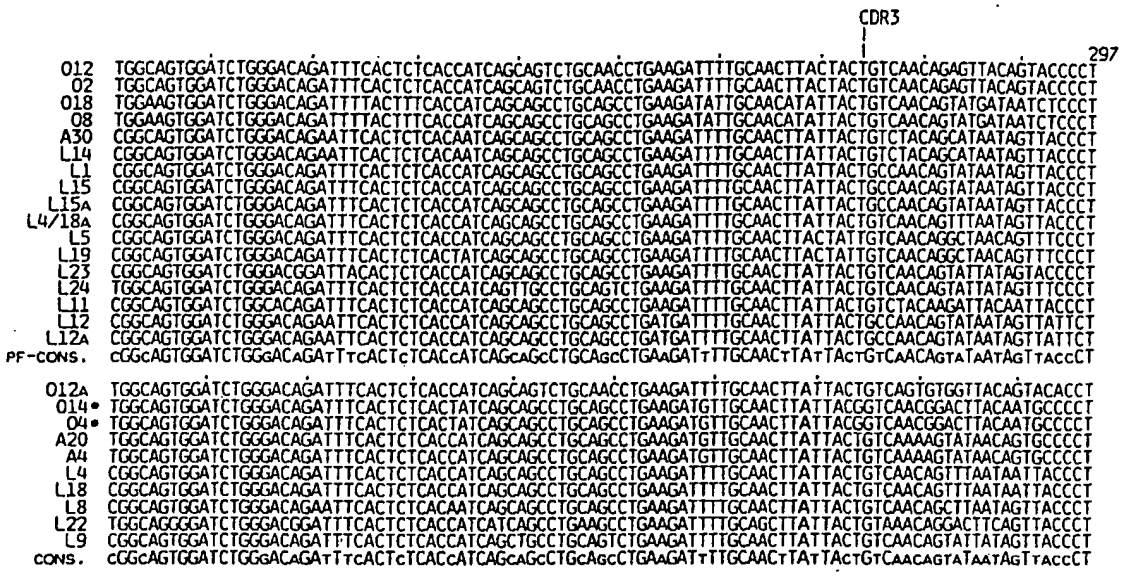


Fig. 2e Continued

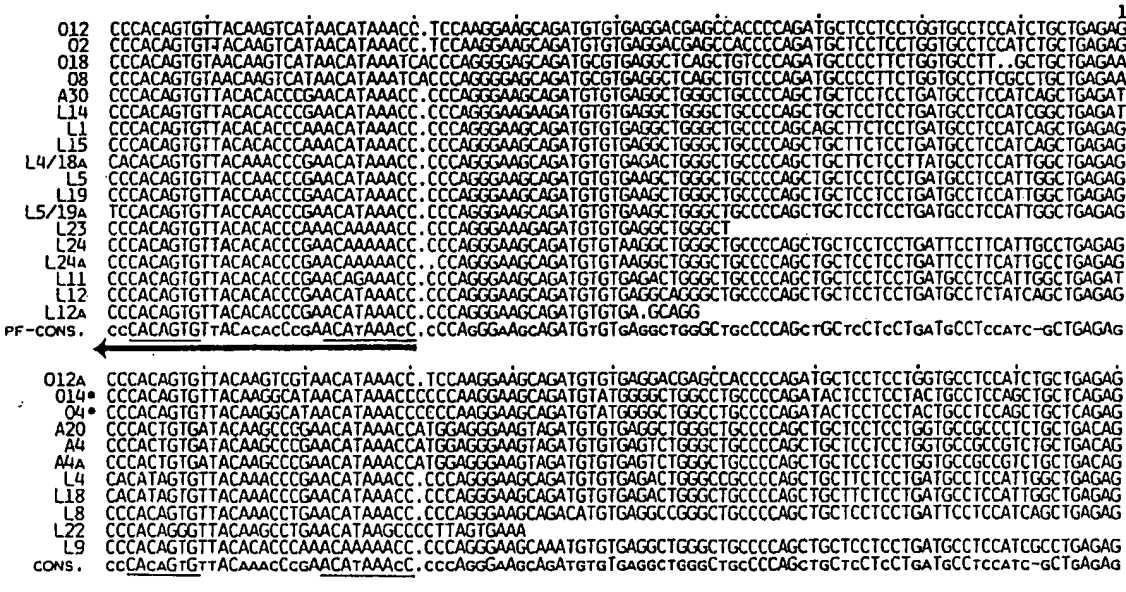


Fig. 2f

```
                                                                  27
O11  CTTTTCCACACCACTGCACCCACCAGG
O1   CTTTTCCACACCACTGCACCCACCAGG
A17  TCTTTCCACCCCACTGCACCCACCAGG
A1   TCTTTCCACCCCACTACACCCACCAGG
A2   TCTTTCCACACCACTGCACCCACCAGG
A19  TCTTTCCACACCACTGCATGCACCAGG
A3   TCTTTCCACACCACTGCATGCACCAGG
A23  TATTTCTACACCACTGCACCCACCAGG
PF-CONS. TCTTTCCACACCACTGCACCCACCAGG

A18  TCTTTCCACACCACTGCACCCACCAGG
A7•  TATTTCTATACCACTGCACCCACCAGG
A21  GCTTAGCAT..CACTGCAC.CACCGGG
A29  TCTTTCCACCCCACTGCACCCACCAGG
A13  TCTTTCCACCCCACTGCACCCACCAGG
L13  TATTTCTTCACTGCTTCACTCACCAGG
CONS. TCTTTCCACACCACTGCACCCACCAGG
```

Fig. 3a

Fig. 3. Alignment of the sequences of the V_κII gene regions.

The presentation is analogous to Fig. 2.

a) The conserved part of the region upstream of dc. Due to many insertions and deletions sequences further upstream are difficult to align. The last nucleotides of this region can be used together with the dc box to construct a primer with V_κII plus V_κIII preference.

b) Sequences upstream of the leader region. One possible primer is shown.

c) Exon 1. For a possible primer sequence see Fig. 2c.

d) 5′ and 3′ parts of the introns. Since the large middle section of the introns is rather divergent an alignment seems not to be meaningful. For a possible primer sequence see Fig. 2d.

e) Exon 2. For primers see Fig. 2e.

f) Where known, the first 100 bp downstream of exon 2 are shown. The hepta- and nonanucleotide recognition sites are underlined. For primers see Fig. 2f.

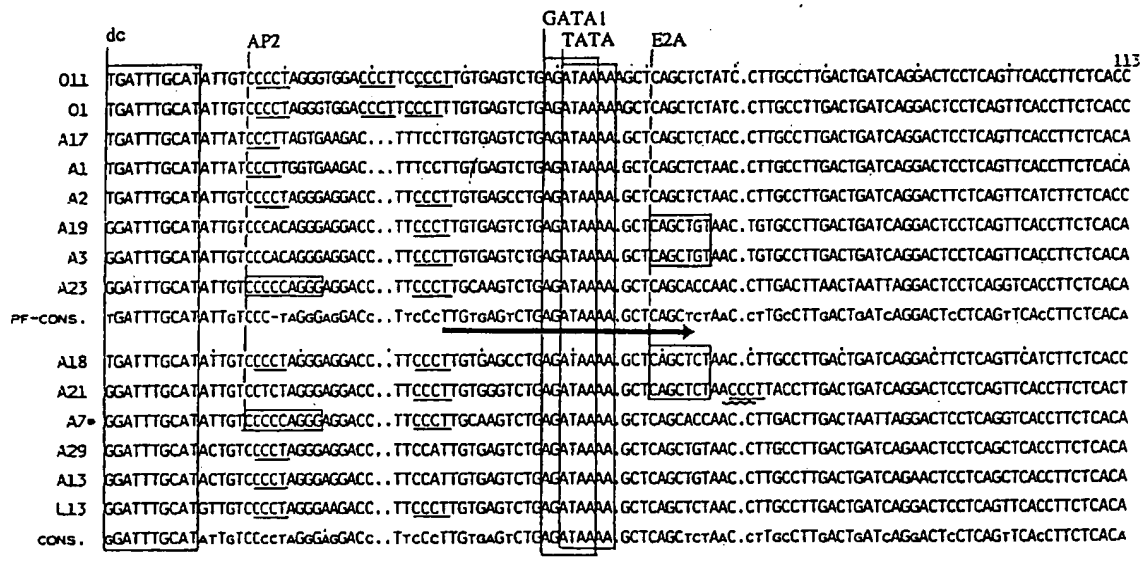


Fig. 3b

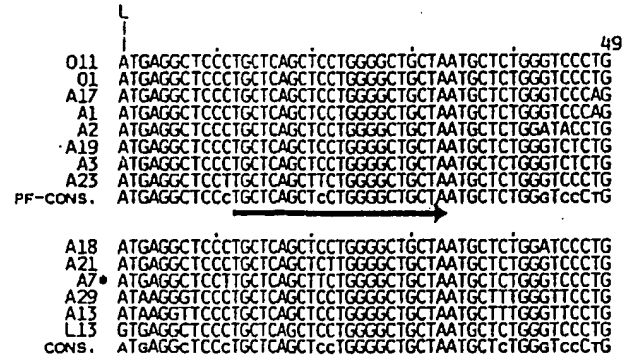


Fig. 3c

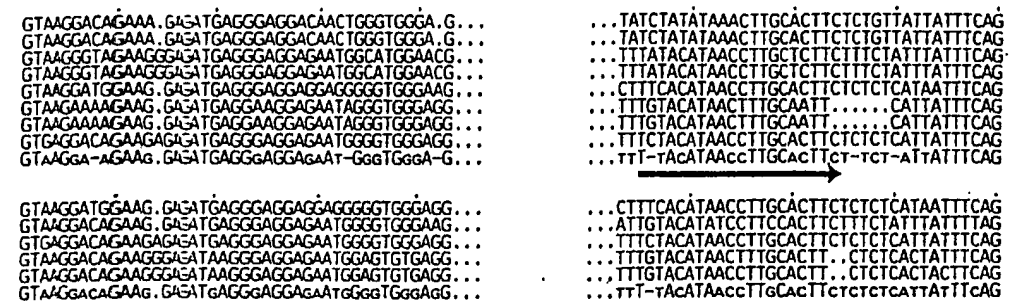


Fig. 3d

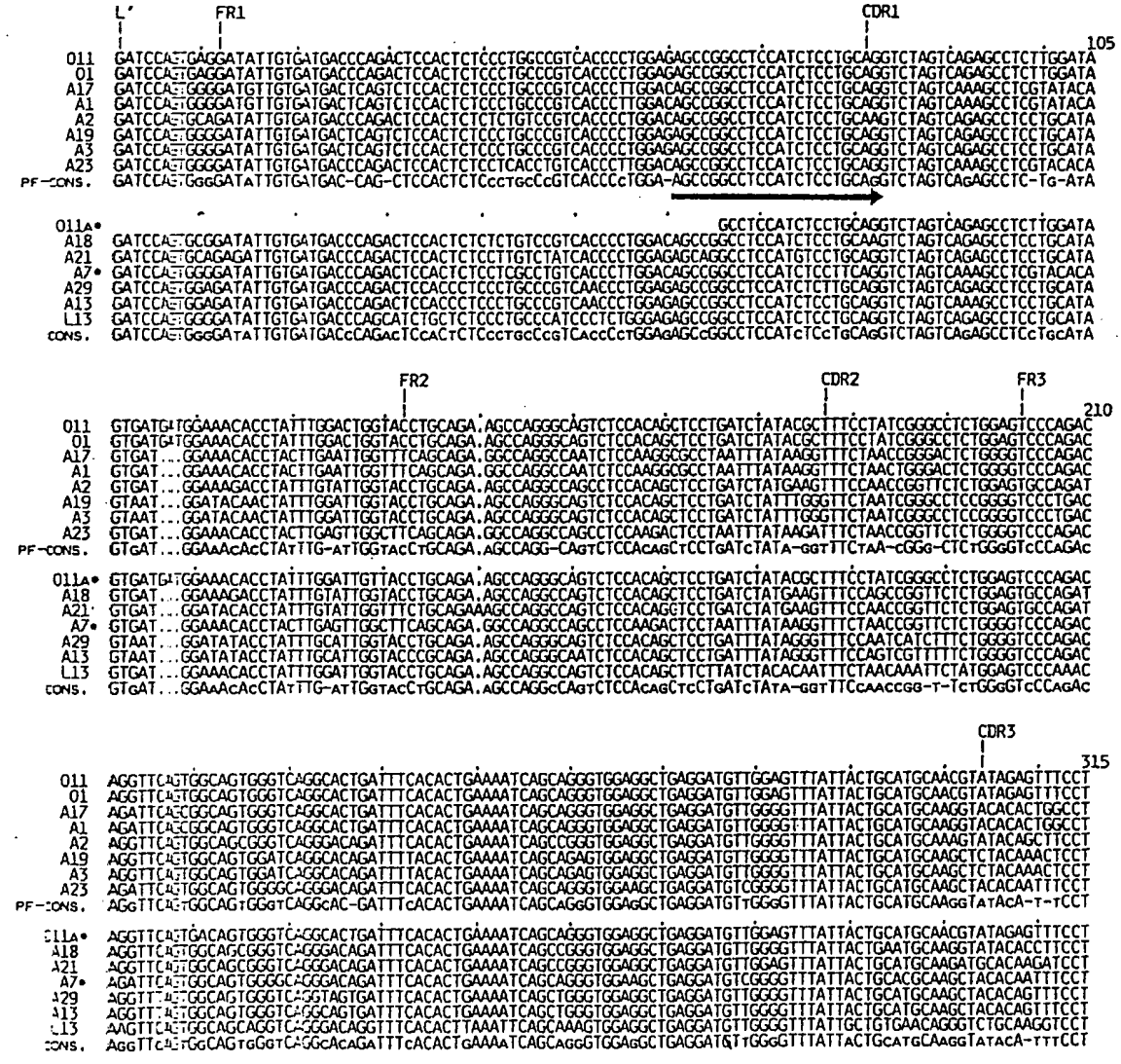


Fig. 3e

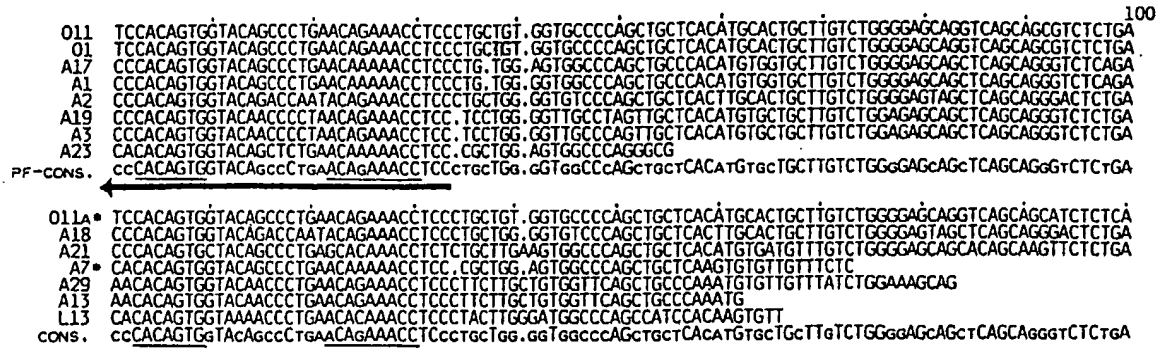


Fig. 3f

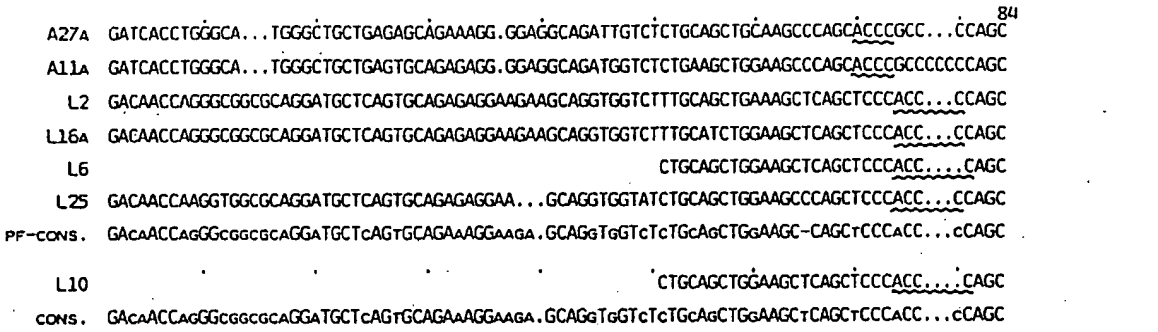


Fig. 4a

Fig. 4. Alignment of the sequences of V_κIII gene regions.

The presentation is analogous to Fig. 2.

a) Sequences upstream of the dc box. For a possible primer see Fig. 3a.

b) Sequences upstream of the leader region. One possible primer is shown.

c) The exon 1 sequences are shown together with 8 bp of upstream sequence in order to include the alternative start codon of L25[12]. For a possible primer see Fig. 2c.

d) Intron. For the possible primer see Fig. 2d.

e) Exon 2. For possible primers see Fig. 2e.

f) 100 bp downstream of exon 2. For primers see Fig. 2f.

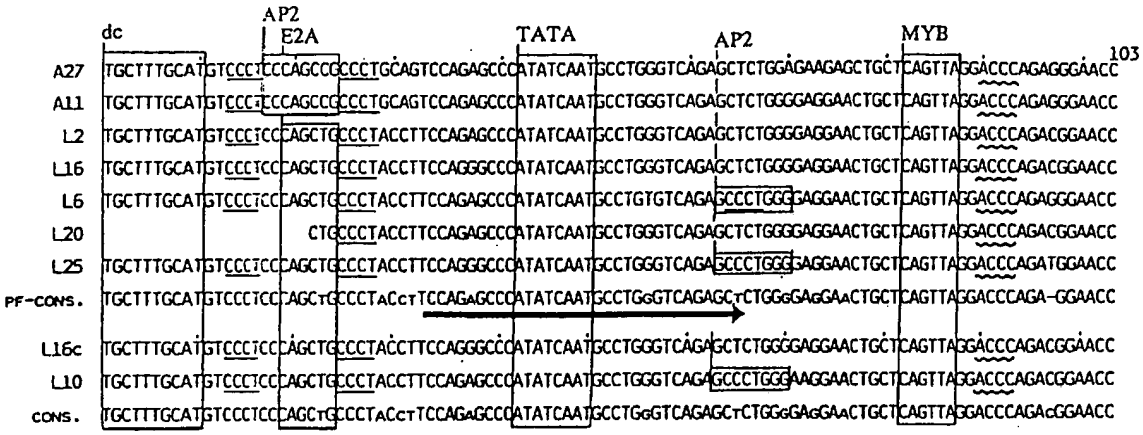


Fig. 4b

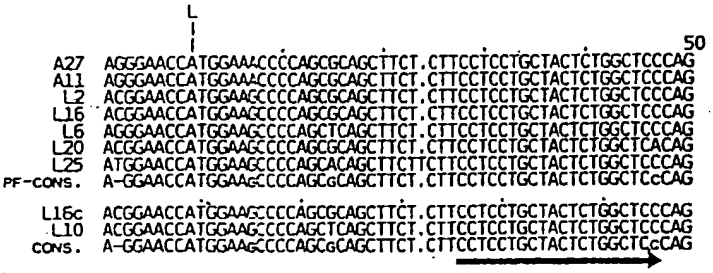


Fig. 4c

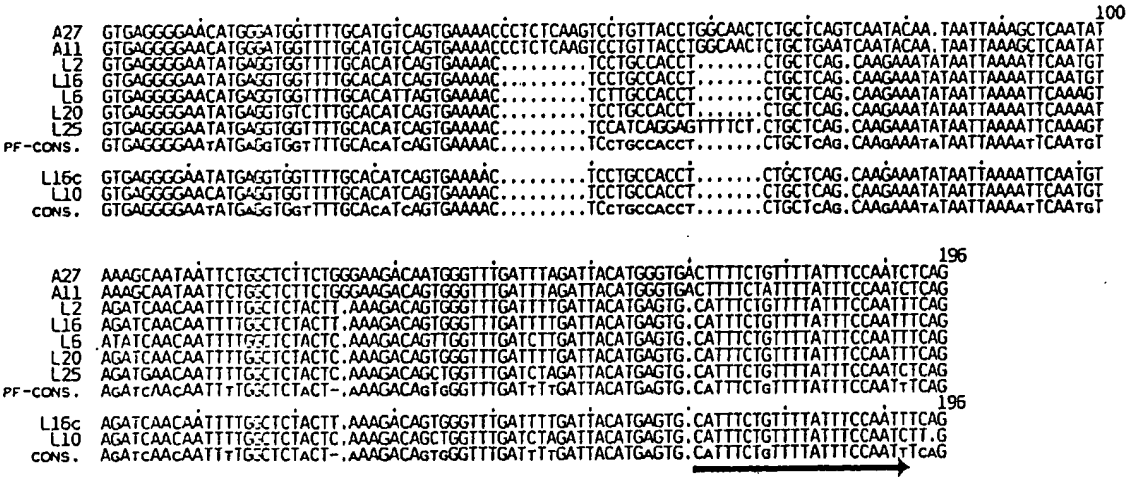


Fig. 4d

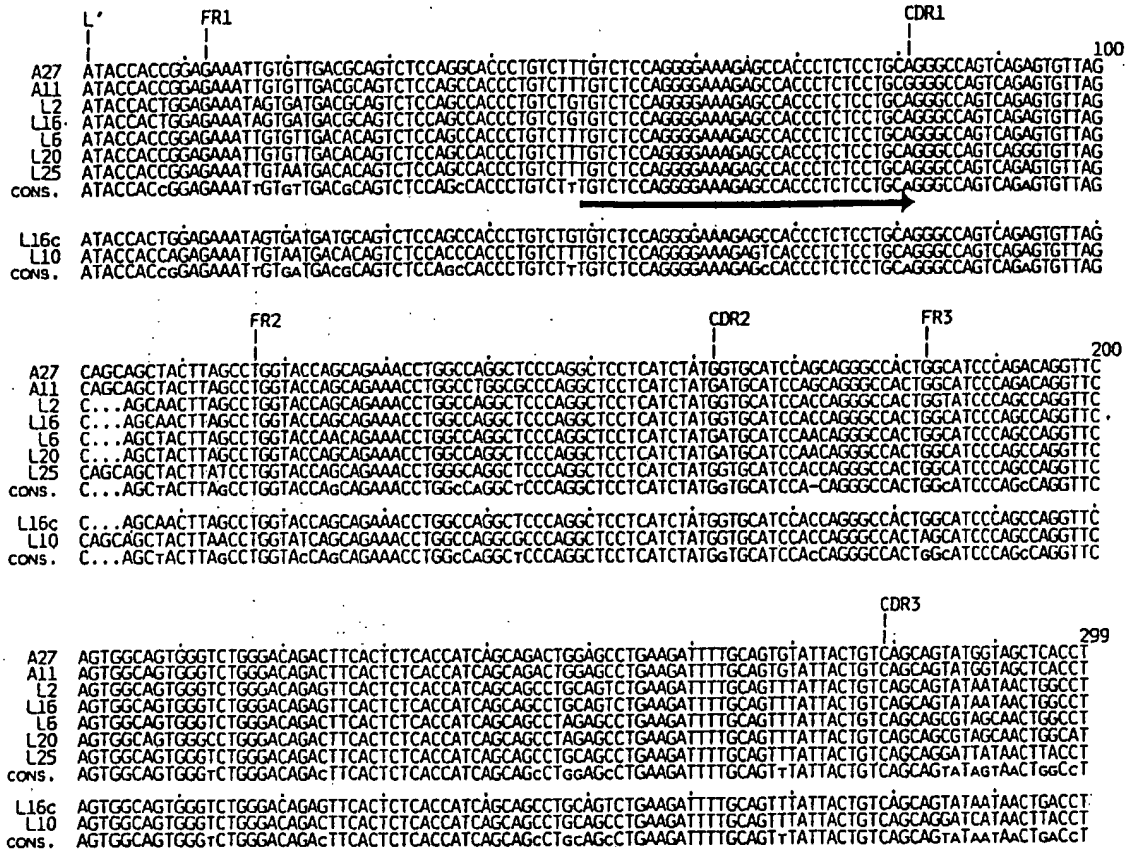

L'
FR1
CDR1
100
A27 ATACCACCGGAGAAATTGTGTTGACGCAGTCTCCAGGCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAG
A11 ATACCACCGGAGAAATTGTGTTGACGCAGTCTCCAGCCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCGGGGCCAGTCAGAGTGTTAG
L2 ATACCACTGGAGAAATAGTGATGACGCAGTCTCCAGCCACCCTGTCTGTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAG
L16 ATACCACTGGAGAAATAGTGATGACGCAGTCTCCAGCCACCCTGTCTGTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAG
L6 ATACCACCGGAGAAATTGTGTTGACACAGTCTCCAGCCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAG
L20 ATACCACCGGAGAAATTGTGTTGACACAGTCTCCAGCCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGGGTGTTAG
L25 ATACCACCGGAGAAATTGTAATGACACAGTCTCCAGCCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAG
CONS. ATACCACcGGAGAAATTGTGTTGACGCAGTCTCCAGcCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAG
L16c ATACCACTGGAGAAATAGTGATGATGCAGTCTCCAGCCACCCTGTCTGTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAG
L10 ATACCACCAGAGAAATTGTAATGACACAGTCTCCACCCACCCTGTCTTTGTCTCCAGGGGAAAGAGTCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAG
CONS. ATACCACCGGAGAAATTGTGATGACGCAGTCTCCAGCCACCCTGTCTTTGTCTCCAGGGGAAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAG
FR2
CDR2
FR3
200
A27 CAGCAGCTACTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGGTGCATCCAGCAGGGCCACTGGCATCCCAGACAGGTTC
A11 CAGCAGCTACTTAGCCTGGTACCAGCAGAAACCTGGCCTGGCGCCCAGGCTCCTCATCTATGATGCATCCAGCAGGGCCACTGGCATCCCAGACAGGTTC
L2 C...AGCAACTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGGTGCATCCACCAGGGCCACTGGTATCCCAGCCAGGTTC
L16 C...AGCAACTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGGTGCATCCACCAGGGCCACTGGCATCCCAGCCAGGTTC
L6 C...AGCTACTTAGCCTGGTACCAACAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGATGCATCCAACAGGGCCACTGGCATCCCAGCCAGGTTC
L20 C...AGCTACTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGATGCATCCAACAGGGCCACTGGCATCCCAGCCAGGTTC
L25 CAGCAGCTACTTATCCTGGTACCAGCAGAAACCTGGGCAGGCTCCCAGGCTCCTCATCTATGGTGCATCCACCAGGGCCACTGGCATCCCAGCCAGGTTC
CONS. C...AGCTACTTAGCCTGGTACCAGCAGAAACCTGGcCAGGCTCCCAGGCTCCTCATCTATGGTGCATCCA-CAGGGCCACTGGcATCCCAGcCAGGTTC
L16c C...AGCAACTTAGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTATGGTGCATCCACCAGGGCCACTGGCATCCCAGCCAGGTTC
L10 CAGCAGCTACTTAACCTGGTATCAGCAGAAACCTGGCCAGGCGCCCAGGCTCCTCATCTATGGTGCATCCACCAGGGCCACTAGCATCCCAGCCAGGTTC
CONS. C...AGCTACTTAGCCTGGTAcCAGCAGAAACCTGGcCAGGCTCCCAGGCTCCTCATCTATGGTGCATCCAcCAGGGCCACTGGcATCCCAGcCAGGTTC
CDR3
299
A27 AGTGGCAGTGGGTCTGGGACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTATTACTGTCAGCAGTATGGTAGCTCACCT
A11 AGTGGCAGTGGGTCTGGGACAGACTTCACTCTCACCATCAGCAGACTGGAGCCTGAAGATTTTGCAGTGTATTACTGTCAGCAGTATGGTAGCTCACCT
L2 AGTGGCAGTGGGTCTGGGACAGAGTTCACTCTCACCATCAGCAGCCTGCAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTATAATAACTGGCCT
L16 AGTGGCAGTGGGTCTGGGACAGAGTTCACTCTCACCATCAGCAGCCTGCAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTATAATAACTGGCCT
L6 AGTGGCAGTGGGTCTGGGACAGACTTCACTCTCACCATCAGCAGCCTAGAGCCTGAAGATTTTGCAGTTTATTACTGTCAGCAGCGTAGCAACTGGCCT
L20 AGTGGCAGTGGGCCTGGGACAGACTTCACTCTCACCATCAGCAGCCTAGAGCCTGAAGATTTTGCAGTTTATTACTGTCAGCAGCGTAGCAACTGGCAT
L25 AGTGGCAGTGGGTCTGGGACAGACTTCACTCTCACCATCAGCAGCCTGCAGCCTGAAGATTTTGCAGTTTATTACTGTCAGCAGGATTATAACTTACCT
CONS. AGTGGCAGTGGGTCTGGGACAGAcTTCACTCTCACCATCAGCAGcCTGgAGcCTGAAGATTTTGCAGTTATTACTGTCAGCAGTATAGTAACTGGCCT
L16c AGTGGCAGTGGGTCTGGGACAGAGTTCACTCTCACCATCAGCAGCCTGCAGTCTGAAGATTTTGCAGTTTATTACTGTCAGCAGTATAATAACTGACCT
L10 AGTGGCAGTGGGTCTGGGACAGACTTCACTCTCACCATCAGCAGCCTGCAGCCTGAAGATTTTGCAGTTTATTACTGTCAGCAGGATCATAACTTACCT
CONS. AGTGGCAGTGGGTCTGGGACAGAcTTCACTCTCACCATCAGCAGcCTGcAGcCTGAAGATTTTGCAGTTATTACTGTCAGCAGTATAATAACTGACCT

Fig. 4e

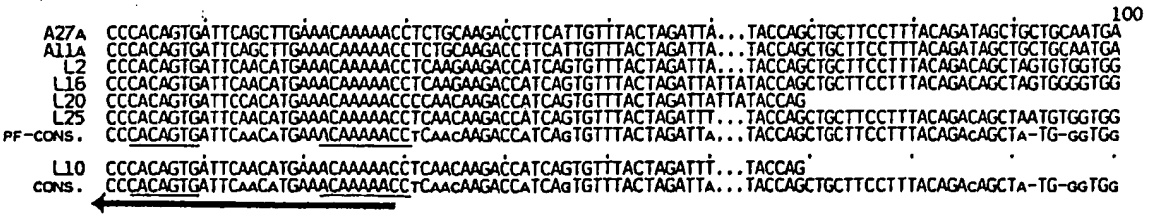

100
A27A CCCACAGTGATTCAGCTTGAAACAAAAACCTCTGCAAGACCTTCATTGTTTACTAGATTA...TACCAGCTGCTTCCTTTACAGATAGCTGCTGCAATGA
A11A CCCACAGTGATTCAGCTTGAAACAAAAACCTCTGCAAGACCTTCATTGTTTACTAGATTA...TACCAGCTGCTTCCTTTACAGATAGCTGCTGCAATGA
L2 CCCACAGTGATTCAACATGAAACAAAAACCTCAAGAAGACCATCAGTGTTTACTAGATTA...TACCAGCTGCTTCCTTTACAGACAGCTAGTGTGGTGG
L16 CCCACAGTGATTCAACATGAAACAAAAACCTCAAGAAGACCATCAGTGTTTACTAGATTATTATACCAGCTGCTTCCTTTACAGACAGCTAGTGGGGTGG
L20 CCCACAGTGATTCCACATGAAACAAAAACCCCAACAAGACCATCAGTGTTTACTAGATTATTATACCAG
L25 CCCACAGTGATTCAACATGAAACAAAAACCTCAACAAGACCATCAGTGTTTACTAGATTT...TACCAGCTGCTTCCTTTACAGACAGCTAATGTGGTGG
PF-CONS. CCCACAGTGATTCAACATGAAACAAAAACCTCAACAAGACCATCAGTGTTTACTAGATTA...TACCAGCTGCTTCCTTTACAGACAGCTA-TG-GGTGG
L10 CCCACAGTGATTCAACATGAAACAAAAACCTCAACAAGACCATCAGTGTTTACTAGATTT...TACCAG
CONS. CCCACAGTGATTCAACATGAAACAAAAACCTCAACAAGACCATCAGTGTTTACTAGATTA...TACCAGCTGCTTCCTTTACAGACAGCTA-TG-GGTGG

Fig. 4f

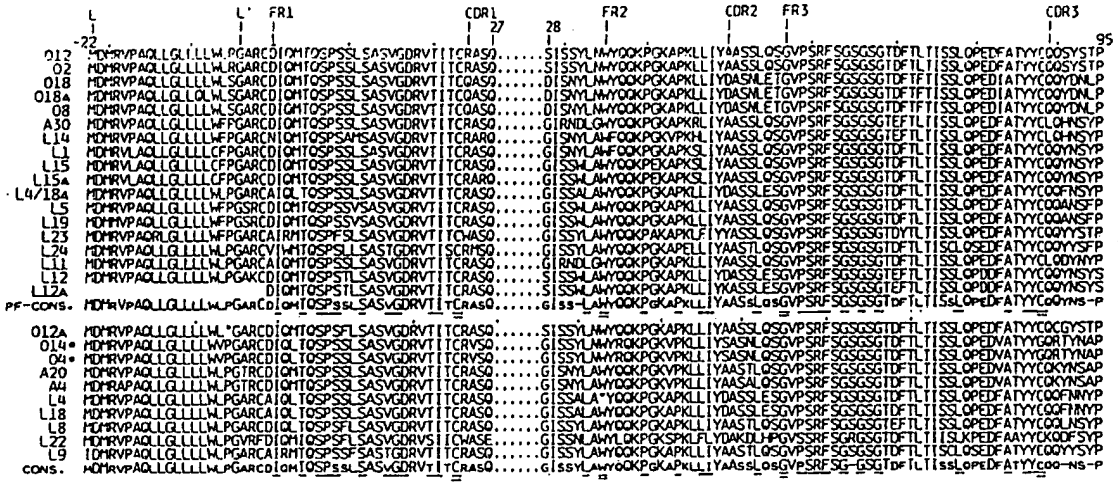


Fig. 5a

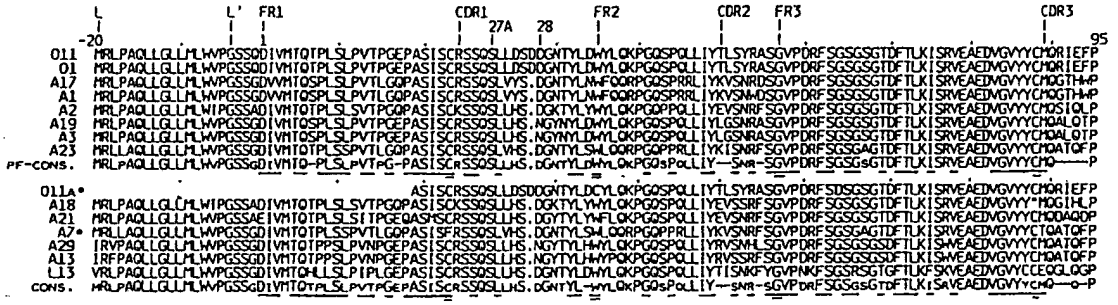


Fig. 5b

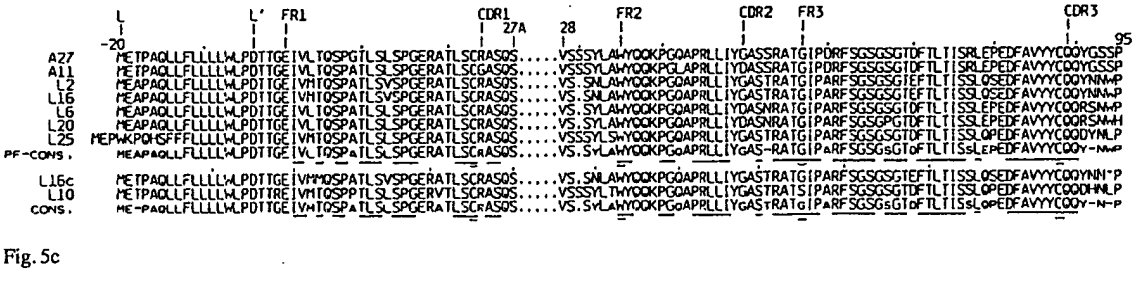


Fig. 5c

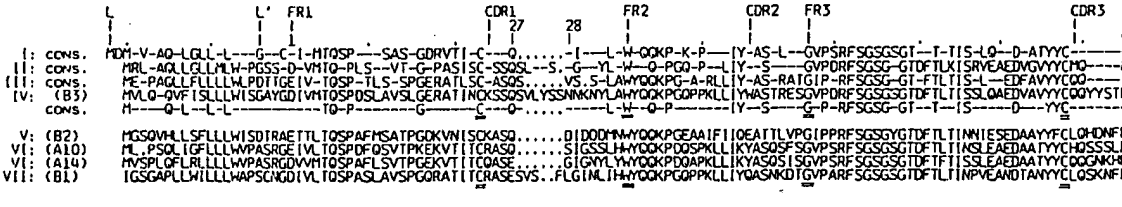


Fig. 5d

Fig. 5. Alignment of the formal translation products of the potentially functional and slightly defective V_x genes of subgroups I–III (a–c).

The numbering of the residues and the definition of invariant residues (underlined) are as in ref.[28]. The four highly conserved residues (see text) are underlined twice. In panel (d) the residues conserved in the sequences of the formal translation products of subgroup I–IV genes are shown.

The work was supported throughout by the *Bundesministerium für Forschung und Technologie* (Center Grant 0316200A) and the *Fonds der Chemischen Industrie*.

References

1 Bentley, D.L. & Rabbitts, T.H. (1980) *Nature (London)* **288**, 730–733.
2 Bentley, D.L. & Rabbitts, T.H. (1983) *Cell* **32**, 181–189.
3 Zachau, H.G., Pech, M., Klobeck, H.-G., Pohlenz, H.-D., Straubinger, B. & Falkner, F.G. (1984) *Hoppe-Seyler's Z. Physiol. Chem.* **365**, 1363–1373.
4 Zachau, H.G. (1989) in *The Immunoglobulin Genes* (Honjo, T., Alt, F.W. & Rabbitts, T.H. eds.) pp. 91–109, Academic Press, London/San Diego.
5 Zachau, H.G. (1990) *Biol. Chem. Hoppe-Seyler* **371**, 1–6.
6 Zachau, H.G. (1993) *Gene*. in press.
7 Weichhold, G.M., Ohnheiser, R. & Zachau, H.G. (1993) *Genomics* **16**, 503–511.
8 Weichhold, G.M., Huber, C., Parnes, J.R. & Zachau, H.G. (1993) *Genomics* **16**, 512–514.
9 Pargent, W., Meindl, A., Thiebe, R., Mitzel, S. & Zachau, H.G. (1991) *Eur. J. Immunol.* **21**, 1821–1827.
10 Lautner-Rieske, A., Huber, C., Meindl, A., Pargent, W., Schäble, K.F., Thiebe, R., Zocher, I. & Zachau, H.G. (1992) *Eur. J. Immunol.* **22**, 1023–1029.
11 Huber, C., Huber, E., Lautner-Rieske, A., Schäble, K.F. & Zachau, H.G. (1993) *Eur. J. Immunol.*, in press.
12 Huber, C., Schäble, K.F., Huber, E., Klein, R., Meindl, A., Thiebe, R., Lamm, R. & Zachau, H.G. (1993) *Eur. J. Immunol.*, in press.
13 Lorenz, W., Schäble, K.F., Thiebe, R., Stavnezer, J. & Zachau, H.G. (1988) *Mol. Immunol.* **25**, 479–484.
14 Klobeck, H.-G. & Zachau, H.G. (1986) *Nucl. Acids Res.* **14**, 4591–4603.
15 Klobeck, H.-G., Zimmer, F.-J., Combriato, G. & Zachau, H.G. (1987) *Nucl. Acids Res.* **15**, 9655–9665.
16 Klein, R., Jaenichen, R. & Zachau, H.G. (1993) in press.
17 Schäble, K.F., Thiebe, R., Meindl, A., Flügel, A. & Zachau, H.G. (1993) in preparation.
18 Huber, C., Thiebe, R., Hameister, H., Smola, H., Lötscher, E. & Zachau, H.G. (1990) *Nucl. Acids Res.* **18**, 3475–3478.
19 Huber, C. (1993) *Doctoral Thesis*, Fakultät für Chemie und Pharmazie, Universität München.
20 Röschenthaler, F., Schäble, K.F., Thiebe, R. & Zachau, H.G. (1992) *Biol. Chem. Hoppe-Seyler* **373**, 177–186.
21 Meindl, A., Klobeck, H.-G., Ohnheiser, R. & Zachau, H.G. (1990) *Eur. J. Immunol.* **20**, 1855–1863.
22 Wilson, A.C., Ochman, H. & Prager, E.M. (1987) *Trends Genet.* **3**, 241–247.
23 Weichhold, G.M., Klobeck, H.-G., Ohnheiser, R., Combriato, G. & Zachau, H.G. (1990) *Nature (London)* **347**, 90–92.
24 Pargent, W., Schäble, K.F. & Zachau, H.G. (1991) *Eur. J. Immunol.* **21**, 1829–1835.
25 Scott, M.G., Crimmins, D.L., McCourt, D.W., Chung, G., Schäble, K.F., Thiebe, R., Quenzel, E.-M., Zachau, H.G. & Nahm, M.H. (1991) *J. Immunol.* **147**, 4007–4013.
26 Schäble, K.F. (1993) *Doctoral Thesis*, Fakultät für Chemie und Pharmazie, Universität München.
27 Straubinger, B., Thiebe, R., Huber, C., Osterholzer, E. & Zachau, H.G. (1988) *Biol. Chem. Hoppe-Seyler* **369**, 601–607.
28 Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S. & Foeller, C. (1991) *NIH Publication* 91-3242, Bethesda, Maryland, USA.
29 Larrick, J.W., Danielson, L., Brenner, C.A., Wallace, E.F., Abrahamson, M., Fry, K.E. & Borrebaeck, C.A.K. (1989) *Bio/Technology* **7**, 934–938.
30 Mullinax, R.L., Gross, E.A., Amberg, J.R., Hay, B.N., Hogrefe, H.H., Kubitz, M.M., Greener, A., Alting-Mees, M., Ardourel, D., Short, J.M., Sorge, J.A. & Shopes, B. (1990) *Proc. Nat. Acad. Sci. USA* **87**, 8095–8099.
31 Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffith, A.D. & Winter, G. (1991) *J. Mol. Biol.* **222**, 581–597.
32 Marks, J.D., Tristem, M., Karpas, A. & Winter, G. (1991) *Eur. J. Immunol.* **21**, 985–991.
33 Kurth, J.H., Mountain, J.L. & Cavalli-Sforza, L.L. (1993) *Genomics* **16**, 69–77.
34 Falkner, F.G. & Zachau, H.G. (1984) *Nature (London)* **310**, 71–74.
35 Parslow, T.G., Blair, D.L., Murphy, W.J. & Granner, D.K. (1984) *Proc. Nat. Acad. Sci. USA* **81**, 2650–2654.
36 Falkner, F.G., Mocikat, R. & Zachau, H.G. (1986) *Nucl. Acids. Res.* **14**, 8819–8827.
37 Mocikat, R., Falkner, F.G. & Zachau, H.G. (1989) in *Tissue Specific Gene Expression* (Renkawitz, R. ed.) pp. 73–85, VCH Verlag, Weinheim.
38 Eckhardt, L.A. (1992) *FASEB J.* **6**, 2553–2560.
39 Wirth, T., Staudt, L. & Baltimore, D. (1987) *Nature (London)* **329**, 174–178.
40 Bergman, Y., Rice, D., Grosschedl, R. & Baltimore, D. (1984) *Proc. Nat. Acad. Sci. USA* **81**, 7041–7045.
41 Högbom, E., Magnusson, A.-C. & Leanderson, T. (1991) *Nucleic Acids Res.* **19**, 4347–4354.
42 Mocikat, R., Pruijn, G.J.M., van der Vliet, P.E. & Zachau, H.G. (1988) *Nucl. Acids. Res.* **16**, 3693–3704.
43 Faisst, S. & Meyer, S. (1992) *Nucleic Acids Res.* **20**, 3–26.
44 Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
45 Hesse, J.E., Lieber, M.R., Mizuuchi, K. & Gellert, M. (1989) *Genes Developm.* **3**, 1053–1061.
46 Milstein, C. (1967) *Nature (London)* **216**, 330–332.
47 Hilschmann, N. (1969) *Naturwissenschaften* **56**, 195–205.
48 Kroemer, G., Helmberg, A., Bernot, A., Auffray, C. & Kofler, R. (1991) *Immunogenetics* **33**, 42–49.
49 Dul, J.L. & Argon, Y. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8135–8139.
50 Zachau, H.G. (1989) in *Evolutionary Tinkering in Gene Expression* (Grunberg-Manago, M., Clark, B.F.C. & Zachau, H.G. eds.) pp. 111–119, Plenum Publishing, London.
51 Straubinger, B., Pech, M., Mühlebach, K., Jaenichen, H.-R., Bauer, H.-G. & Zachau, H.G. (1984) *Nucl. Acids. Res.* **12**, 5265–5275.
52 Manheimer-Lory, A., Katz, J.B., Pillinger, M., Ghossein, C., Smith, A. & Diamond, B. (1991) *J. Exp. Med.* **174**, 1639–1652.
53 Jaenichen, H.-R., Pech, M., Lindenmaier, W., Wildgruber, N. & Zachau, H.G. (1984) *Nucl. Acids. Res.* **12**, 5249–5263.
54 Chen, P.P., Albrandt, K., Orida, N.K., Radoux, V., Chen, E.Y., Schrantz, R., Liu, F.-T. & Carson, D.A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8318–8322.
55 Scott, M.G., Crimmins, D.L., McCourt, D.W., Zocher, I., Thiebe, R., Zachau, H.G. & Nahm, M.H. (1989) *J. Immunol.* **143**, 4110–4166.
56 Straubinger, B., Huber, E., Lorenz, W., Osterholzer, E., Pargent, W., Pech, M., Pohlenz, H.-D., Zimmer, F.-J. & Zachau, H.G. (1988) *J. Mol. Biol.* **199**, 23–34.
57 Straubinger, B., Osterholzer, E. & Zachau, H.G. (1987) *Nucl. Acids Res.* **15**, 9567–9575.
58 Radoux, V., Chen, P.P., Sorge, J.A. & Carson, D.A. (1986) *J. Exp. Med.* **164**, 2119–2124.

59 Liu, M.-F., Robbins, D.L., Crowley, J.J., Sinha, S., Kozin, F., Kipps, T.J., Carson, D.A. & Chen, P.P. (1989) *J. Immunol.* **142**, 688–694.
60 Pech, M., Jaenichen, H.-R., Pohlenz, H.-D., Neumaier, P.S., Klobeck, H.-G. & Zachau, H.G. (1984) *J. Mol. Biol.* **176**, 189–204.
61 Pech, M., Smola, H., Pohlenz, H.D., Straubinger, B., Gerl, R. & Zachau, H.G. (1985) *J. Mol. Biol.* **183**, 291–299.
62 Pech, M. & Zachau, H.G. (1984) *Nucl. Acids Res.* **12**, 9229–9236.
63 Victor, K.D., Randen, I., Thompson, K., Forre, O., Natvig, J.B., Fu, S.M. & Capra, J.D. (1991) *J. Clin. Invest.* **87**, 1603–1613.
64 Stavnezer, J., Kekish, O., Batter, D., Grenier, J., Balazs, I., Henderson, E. & Zegers, B.J.M. (1985) *Nucl. Acids Res.* **13**, 3495–3514.
65 Klobeck, H.-G., Bornkamm, G.W., Combriato, G., Mocikat, R., Pohlenz, H.-D. & Zachau, H.G. (1985) *Nucl. Acids Res.* **13**, 6515–6529.
66 Pohlenz, H.-D., Straubinger, B., Thiebe, R., Pech, M., Zimmer, F.-J. & Zachau, H.G. (1987) *J. Mol. Biol.* **193**, 241–253.
67 Klopstock, T. (1991) *Doctoral Thesis*, Medizinische Fakultät der Universität München.
68 Zimmer, F.-J., Huber, C., Quenzel, M., Schek, H., Stiller, C., Thiebe, R. & Zachau, H.G. (1990) *Biol. Chem. Hoppe-Seyler* **371**, 283–290.
69 Lötscher, E., Grzeschik, K.-H., Bauer, H.G., Pohlenz, H.-D., Straubinger, B. & Zachau, H.G. (1986) *Nature (London)* **320**, 456–458.
70 Lötscher, E., Zimmer, F.-J., Klopstock, T., Grzeschik, K.-H., Jaenichen, R., Straubinger, B. & Zachau, H.G. (1988) *Gene* **69**, 215–223.
71 Straubinger, B., Thiebe, R., Pech, M. & Zachau, H.G. (1988) *Gene* **69**, 209–214.

Karlheinz F. Schäble and Hans G. Zachau, Institut für Physiologische Chemie, Universität München, Schillerstr. 44, D-80336 München.

Biol.Chem.Hoppe-Seyler 374, 1001-1022 (1993)
Schäble and Zachau

Addendum to Table 1

a) Cox, Tomlinson and Winter amplified from genomic DNA by PCR FR1-CDR3 sequences and called four of them "new gene segments" (Eur. J. Immunol. 24, 827-836, 1994). However, the sequences do not correspond to new gene loci but have to be considered as alleles of published V_κ genes and orphons, as far as this can be concluded in the absence of intron sequences and data on the genomic context.

DPK14, A21/A5; DPK37, Z3; LFVK5, V118; LFVK431, L1.

The comparisons are described in Klein and Zachau, Ann. NY Acad. Sci., in press.

b) The reference for the gene A11a is [54].

c) Alternative accession numbers
for Z2, Z3, and Z4 are S37418/19/21.

Corrigendum of Table 3

Subgroup III: intron length [bp] 168-1446;
similarity [%] 64-100.

Update of the list of references

4 (1994) 2. Edition, in press.
6 (1993) Gene 135, 167-173.
11 (1993) Eur. J. Immunol. 23, 2860-2867.
12 (1993) Eur. J. Immunol. 23, 2868-2875.
16 (1993) Eur. J. Immunol. 23, 3248-3271.
17 (1994) Biol. Chem. Hoppe-Seyler 375, 189-199.
19 (1994) Genomics 22, 213-215.

Dedicated to the memory of Georges Kohler

Immunoglobulin Genes
Second edition

Edited by

T. Honjo
Kyoto University, Japan

F. W. Alt
Howard Hughes Medical Institute Research Laboratories,
The Children's Hospital, Boston

ACADEMIC PRESS

Harcourt Brace & Company, Publishers
London, San Diego, New York, Boston, Sydney, Tokyo, Toronto

ACADEMIC PRESS LIMITED
24–28 Oval Road
LONDON NW1 7DX, UK

U. S. Edition Published by
ACADEMIC PRESS INC.
San Diego, CA 92101, USA

This book is printed on acid-free paper

A catalogue record for this book is available from the British Library

ISBN 0 12 053640 4

Typeset by Phoenix Photosetting, Chatham
Printed and bound by Hartnolls Ltd, Bodmin, Cornwall

Contents

Contributors

FW Alt
Howard Hughes Medical Institute Research Laboratories, The Childr Hospital, 300 Longwood Avenue, Boston, MA 02115, USA

MK Anderson
Department of Molecular Genetics, All Children's Hospital, 801 Sixth St South, St Petersburg, FL 33701-4899, USA

PD Burrows
Division of Developmental and Clinical Immunology, Departments of Mi biology, Medicine and Pediatrics, University of Alabama at Birmingham, Wallace Tumor Institute, UAB Station, Birmingham, AL, USA

K Calame
Columbia University College of Physicians and Surgeons, Departmer Microbiology, 701 West 168th Street, New York, NY 10032-2704, USA

JD Capra
Southwestern Medical School, Department of Microbiology, 5323 H Hines Boulevard, Dallas, TX 75235, USA

MD Cooper
Howard Hughes Medical Institute, Department of Pediatrics Microbiology, University of Alabama at Birmingham, 263 Wallace Tu Institute, UAB Station, Birmingham, AL 35294, USA

LE Daltch
Tufts University School of Medicine, Department of Pathology, 136 Harr Avenue, Boston, MA 02111, USA

EA Faust
Howard Hughes Medical Institute Research Laboratories, Universit California, Los Angeles, 5-748 MacDonald Building, 10833 Le C Avenue, Los Angeles, CA 90024-1662, USA

S Ghosh
Howard Hughes Medical Institute, Yale University, 295 Congress Ave 154 BCMM, New Haven, CT 06510, USA

LA Herzenberg
Stanford University School of Medicine, Department of Genetics, Imm genetics and Cell Sorting Laboratory, Stanford, CA 94305-1662, USA

7

Immunoglobulin heavy chain loci of mouse and human

Tasuku Honjo and Fumihiko Matsuda

Department of Medical Chemistry, Faculty of Medicine and Center for Molecular Biology and Genetics, Kyoto University, Sakyo-ku, Kyoto 606, Japan

INTRODUCTION

The immunoglobulin (Ig) molecule is composed of the heavy (H) and light (L) chains, both of which consist of the variable (V) and constant (C) regions. The V region is responsible for antigen binding whereas the C_H region specifies the isotype of Ig. Genes encoding IgH V regions are split into V_H, D_H and J_H segments, each of which is comprised of multiple copies. One each of the three segments is generally assembled into a functional V_H gene by a somatic genetic event called VDJ recombination. The V_H locus that contains these gene segments is located on human chromosome 14q32.33 (Croce *et al.*, 1979; Kirsch *et al.*, 1982) or mouse chromosome 12 (D'eustachio *et al.*, 1980). There are nine and eight C_H genes in the human and murine C_H loci, respectively. The cluster of C_H genes is referred to as the C_H locus. The V_H and C_H loci are tightly linked on the chromosome. The distance between the V_H locus segment nearest the 3′ end (J_H) and the C_H gene nearest the 5′ end (C_μ) is less than 8 kb in both mouse and human (Liu *et al.*, 1980; Ravetch *et al.*, 1981). Recent studies have shown that the human V_H and C_H loci span at least 1.1 megabases (Mb) (Matsuda *et al.*, 1993; Cook *et al.*, 1994) and 0.3 Mb (Bottaro *et al.*, 1989; Hofker *et al.*, 1989), respectively. Thus the IgH locus combining the V_H and C_H loci constitutes a huge multigene family. The aim of this chapter is to summarize the current knowledge of the organization and structure of the mouse and human IgH loci and to discuss its biological significance and implications.

IMMUNOGLOBULIN GENES (2ND EDITION)
ISBN 0-12-053640-4

There are a number of reasons why the complete elucidation of the IgH locus is important to immunologists as well as geneticists. From a geneticist's point of view it is fascinating to elucidate how the multigene family has evolved and how a large number of V_H segments are maintained. These questions may be answered more clearly by direct comparison of the IgH loci between related species. Complete knowledge of the organization and structure of the germline V_H segments may provide answers to a number of questions essential to Ig repertoire formation and its expression. Obviously, the total number of V_H segments determines the upper limit of the germline Ig repertoire, although somatic genetic events including VDJ recombination and hypermutation amplify the expressed repertoire tremendously. Immunologists are interested in polymorphic variation of the number and repertoire of germline V_H segments and C_H genes and in the association of such polymorphisms with disease susceptibility. It is known that some V_H segments are overrepresented, suggesting that V_H usage may not be random. It is important to know whether the germline organization or structure of V_H segments have anything to do with such biased usage. Isolation of the total human V_H segments has played important roles for generation of human Ig in J segment-disrupted mice carrying human Ig mini loci (Green *et al.*, 1994; Taylor *et al.*, 1994). Needless to say, studies on the complete organization of the C_H locus are the basis for understanding the molecular mechanism of class switching (see Chapter 11).

Since the publication of the first edition of this book (1989), a major breakthrough was made in elucidation of the human IgH locus by completion of the physical mapping of human V_H segments using yeast artificial chromosome (YAC) clones. Organization and structure of human V_H segments have been extensively studied, providing a tremendously useful reference to map expressed V_H genes, and their polymorphisms. Unfortunately, however, little progress has been made in the study of the mouse V_H locus, though comparison of V_H organization between mouse and human would be extremely interesting.

HUMAN V_H LOCUS

V_H subgroups and families

Human V_H regions were divided into three subgroups based on amino acid sequences (reviewed in Kabat *et al.*, 1991). These protein subgroups have been further subdivided into six distinct V_H families defined by nucleotide sequence homology. V_H segments that show 80% or greater similarity are considered to be in the same family, while V_H segments that have less than 70% similarity to one another form different V_H families (Kodaira *et al.*, 1986; Lee *et al.*, 1987; Shen *et al.*, 1987; Berman *et al.*, 1988). Such criteria have been supported by construction of the phylogenetic tree of 33 functional V_H segments located in the 3′ 0.8-Mb region of the V_H locus (Fig. 1) (Haino *et al.*, 1994). According to this phylogenetic tree human V_H segments first

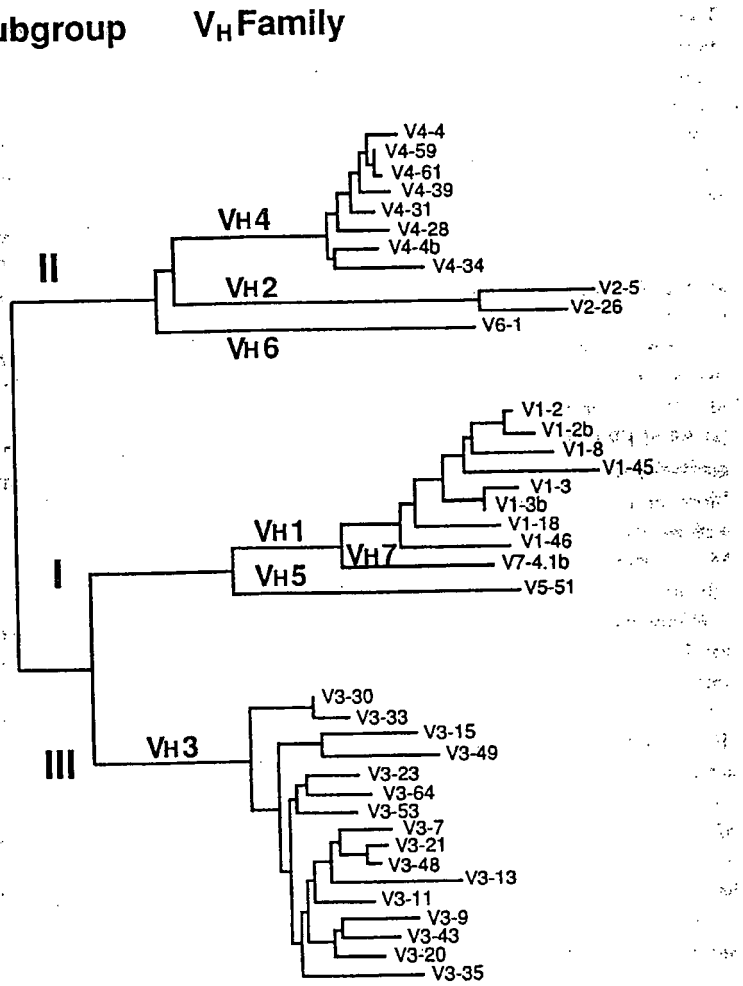


Fig. 1. Phylogenetic tree of human germline functional V_H segments. (Modified from Haino *et al.*, 1994.)

diverged into two groups: subgroup II and ancestor of subgroup I and III. Subgroup II was then divided into V_H2, V_H4 and V_H6 families. Subgroup I was split into V_H1 and V_H5 families. A unique set of V_H segments, which share high homology (78–82%) with V_H1 but differ from V_H1 at a clustered region between framework 2 (FR2) and FR3, has been proposed to be classified as V_H7 family (Schroeder *et al.*, 1990). According to the above definition of the V_H family, V_H7 should be a subfamily of V_H1 or a family captured in transition from V_H1 to independence (Kirkham and

Schroeder, 1994). Nonetheless, the classification of V_H7 is useful as at least six members of V_H7 have been found and mapped at dispersed positions in the V_H locus (van Dijk *et al.*, 1993). Subgroup III contains the largest number of members, yet constitutes a single family, V_H3. It is interesting to note that the V_H4 (Lee *et al.*, 1987), V_H5 (Shen *et al.*, 1987), V_H6 (Berman *et al.*, 1988) and V_H7 families have been identified by comparison of nucleotide sequences of V_H segments. The V_H4 family members are most strongly conserved, suggesting that V_H4 may have evolved most recently (Lee *et al.*, 1987; Haino *et al.*, 1994). However, frequent recombination between V_H segments makes it difficult to estimate the precise time of divergence among V_H segments. The V_H5 and V_H6 families contain only two and one members, respectively.

There are several V_H family-specific conserved regions in human germline V_H segments (Kabat *et al.*, 1991; Tomlinson *et al.*, 1992; Matsuda *et al.*, 1993; Haino *et al.*, 1994). Family-specific sequences were found in codons 9–30 in FR1 and codons 60–85 of FR3. It is important to note that codons 60–65 in the 3′ portion of complementarity determining region 2 (CDR2) were conserved in a family-specific way. More or less universally conserved were codons 1–8, FR2 (codons 38–47) and codons 86–92, in which the embedded heptamer recombination signal is located. More extensive structural comparison of V_H subregions is found elsewhere (Tomlinson *et al.*, 1992; Kirkham and Schroeder, 1994).

Comparison of upstream sequences of V_H segments revealed striking family-specific conservation. The locations of the octamer motif and TATA box are different among families (Haino *et al.*, 1994). A heptamer sequence with consensus CTCATGA is located 2–22 bp upstream of the octamer motif in mouse V_H segments (Eaton and Calame, 1987; Siu *et al.*, 1987) and is required for full V_H promoter activity in mouse lymphoid cells (Ballard and Bothwell, 1986; Eaton and Calame, 1987). Although the heptamer element is located 2 bp upstream of the octamer motif of the human V_H1 family, the heptamer element is not detectable around similar places of the other families. This finding does not support the hypothesis that the heptamer element is involved in the activation of the H-chain promoter by the *oct* protein before the activation of the L-chain promoter, which does not contain the heptamer motif (Kemler *et al.*, 1989).

Physical mapping of human V_H segments

Studies on physical mapping of the human V_H locus were initiated by cosmid cloning (Kodaira *et al.*, 1986). Distribution of V_H families on 23 cosmid clones with average size 40 kb has shown that members of different V_H families are interspersed, in contrast to the finding that the same family members tend to cluster in the mouse V_H locus (Kemp *et al.*, 1981; Rechavi *et al.*, 1982). Another important conclusion from early studies on physical mapping using phage and cosmid vectors is the presence of abundant pseudogenes (about 40%), many of which are highly conserved with only a few point mutations (Givol *et al.*, 1981; Kodaira *et al.*, 1986). A similar type of

analysis mapped the D3 segment only 22 kb upstream of the J_H cluster (Buluwela *et al.*, 1988; Matsuda *et al.*, 1988) and the V_H6 segment 20 kb upstream to the D4 segment (Buluwela and Rabbitts, 1988; Sato *et al.*, 1988; Schroeder *et al.*, 1988).

A more general overview of the whole human V_H locus has been provided by studies using pulsed field gel electrophoresis (PFG). Human DNA digested with rare restriction site enzymes was separated by PFG and hybridized with various V_H family-specific probes. Such an analysis allowed the V_H content to be examined on DNA fragments of a few hundred to one thousand kilobases. The total size of the human V_H locus was estimated to be about 2.5–3.0 Mb (Berman *et al.*, 1988; Matsuda *et al.*, 1988) including the D5-hybridizing fragments that later mapped to chromosome 15. Such studies also confirmed the previous conclusion that human V_H families are intermingled. PFG analysis using two-dimensional electrophoresis has provided a more precise determination of the total V_H locus of about 1.2 Mb, on which 76 human V_H segments were mapped using *Sfi*I, *Bss*HII and *Not*I digests (Walter *et al.*, 1990). Although the precise location of each V_H segment cannot be determined and some V_H segments were inevitably missed, this study has made a great contribution to defining the overall organization of the human V_H locus. The same group further refined the mapping using the deletion profile of V_H segments associated with VDJ recombination in human B cell lines (Walter *et al.*, 1991a).

Introduction of the YAC vector has been essential to complete the physical mapping of the human V_H locus (Fig. 2). The first report using YAC cloning identified and located five V_H segments proximal to the D_H segments (Shin *et al.*, 1991). These authors proposed to rename all the V_H segments by the family number and the order from the 3′ end of the V_H locus. For example, V3-36P indicates a V_H3 family member located thirty-sixth from the V_H segment nearest the 3′ end, i.e. V6–1. P indicates the pseudogene. An insertional polymorphic V_H segment is indicated by a number with decimal point. This nomenclature of V_H segments was controversial not only because the investigators named V_H segments idiosyncratically but also because many expressed V_H sequences containing somatic mutations could not be easily assigned as different V_H segments. The newly proposed nomenclature defined V_H segments only when they were mapped on the chromosome, which had been expected to be completed in a few years. The same group in Kyoto has completed mapping of about 70% (0.8 Mb) of the human V_H locus by analysing more than seven overlapping YAC clones. All the YAC clones were subcloned into either cosmids, phages or plasmids, and the nucleotide sequences of 64 V_H segments were determined (Matsuda *et al.*, 1993). V_H segments at around 770–740, 710, 555, 430, 360 and 200–100 kb upstream of the J_H segments have the same transcriptional orientation as the J_H segments. The results indicate that there is no gross inversion in the human V_H locus, and that the majority of V_H segments rearrange to associate with the D_H and J_H segments by looping-out but not by inversion, in contrast to the human V_κ locus (see Chapter 8).

Subsequently, Cook *et al.* (1994) identified the 5′ end of the human V_H locus using human telomere activity in yeast and a chromosome translocation that places telomere-proximal V_H segments onto chromosome 8. A 200-kb clone (yIgH6) was isolated and subcloned into cosmids. V_H family-specific primers were used to

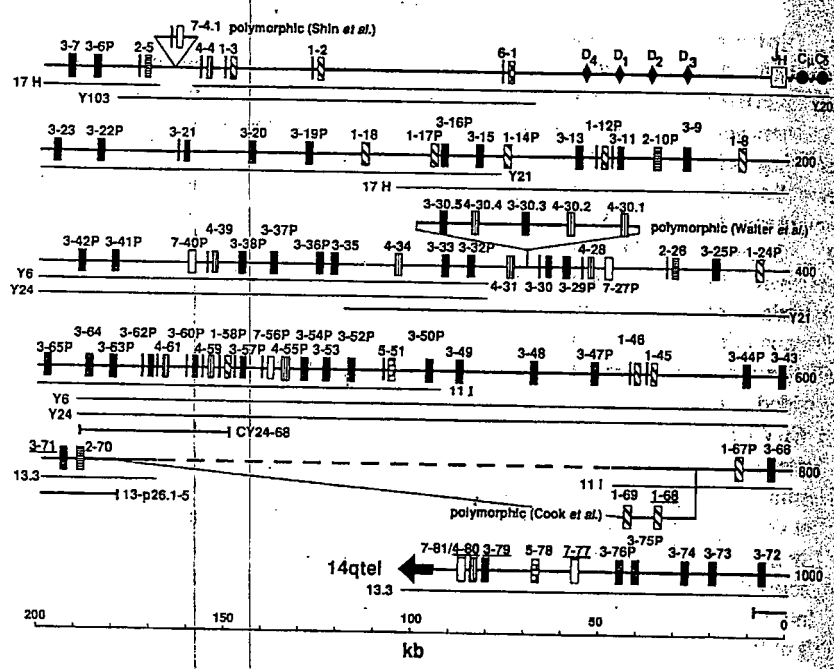


Fig. 2. Organization of the human IgH locus. The 1.1-Mb DNA is shown by six thick horizontal lines with the 3′ end at the top right corner. YAC and cosmid clones covering the region are shown with their names by horizontal lines below. V_H segments belonging to V_H1, 2, 3, 4, 5, 6 and 7 families are indicated by the symbols ▨, ▤, ■, ▥, ▧, ▦ and □. Five V_H segments whose sequences have not been determined are underlined.

amplify V_H coding regions and to determine 19 V_H sequences identified on yIgH6. The two V_H segments nearest the 3′ end on yIgH6 were virtually identical to V3-63P and V3-64, which are the V_H segments nearest the 5′ end previously identified by the Kyoto group (Matsuda *et al.*, 1993), thus linking the entire V_H locus physically by YAC clones. The Kyoto group also isolated an independent clone containing the V_H segments nearest the 5′ end and the telomere (F. Matsuda *et al.*, unpublished data). Interestingly, the distance between the telomere sequence and the 5′-end V_H segment (V7-81) is only a few kilobases. The telomere-containing YAC clone (13.3) of the Kyoto group completely overlaps with yIgH6 down to V2-70, confirming the telomeric end of the physical map by the Cambridge group. However, another YAC clone (11I) extending toward the telomeric end from V3-50P contains a striking difference starting from V1-67P. The Kyoto group carried out two-colour *in situ* hybridization of interphase nuclei using cosmid clones as probes and estimated that the gap between CY24-68 and 13-p26.1-5 (Fig. 2) is either 194 ± 27 kb or 109 kb depending on haplotypes. The distance between the same clones is about 90 kb

Table I Summary of the human V_H segments*

	V_H families							
Chromosomes (length)	1	2	3	4	5	6	7	Total
14q32 (1100 kb)	14	4	48	12	2	1	6	87
15q11 (>250 kb)	6	0	1	1	0	0	0	8
16p11 (>700 kb)	4	1	11	0	0	0	0	16
Total (≤2 Mb)	24	5	60	13	2	1	6	111

* The number of V_H segments on chromosome 14 is calculated by the results from Matsuda *et al.* (1993) and Cook *et al.* (1994). Seven polymorphic V_H segments reported (Shin *et al.*, 1991; Walter *et al.*, 1993) are included. Information of V_H segments on chromosomes 15 and 16 is taken from Nagaoka *et al.* (1994) and Tomlinson *et al.* (1994).

according to the map of yIgH6. It is clear that there is a large insertion polymorphism (about 90 kb) between V1-67P and V2-70. yIgH6 is likely to represent a haplotype with a large deletion as described (Walter *et al.*, 1990). The physical mapping of the human V_H locus completed by the linkage between yIgH6 and Y6 (and Y24) shows that the total number of V_H segments is 87, although there are several polymorphic insertions or deletions (Table I). The only caveat to the above conclusion is cloning artefacts by YAC cloning. The best way to exclude such artefacts is to isolate independent overlapping clones, which has been done for most of the V_H regions so far analysed.

V_H segment number and polymorphisms

One of the most important goals in the study of the human V_H locus is to determine the total number of V_H segments. Given the almost complete physical map of the human V_H locus, the total number of V_H segments in the smallest haplotype is calculated to be 81 with an additional six or more polymorphic V_H segments (Table I), which is in general agreement with the previous estimation using two-dimensional PFG (Walter *et al.*, 1990) or specific amplification of V_H segments by polymerase chain reaction (PCR) (Tomlinson *et al.*, 1992). Polymorphic insertion (or deletion) appears to cluster at three regions: V4-4/V2-5, V3-30/V4-31 and V1-67P/V2-70. It is interesting to know whether any particular DNA sequences are responsible for these clustered polymorphic deletions (or insertions). A project to determine the entire nucleotide sequence of the human V_H locus is under way by the Kyoto group and will provide an answer to the above question. As previously analysed there are abundant pseudogenes. At least 32 out of 87 mapped V_H segments are pseudogenes; four V_H segments were not found in the expressed V_H database; 46 V_H segments were shown to be used; and the remaining five V_H segments were unsequenced (Matsuda *et al.*, 1993; Cook *et al.*, 1994; Haino *et al.*, 1994). The 46 V_H segments used are classified into 8 V_H1, 3 V_H2, 22 V_H3, 10 V_H4, 1 V_H5, 1 V_H6 and 1 V_H7 family segments.

Restriction fragment length polymorphism (RFLP) and DNA sequencing have shown that there are a number of polymorphic V_H alleles. One of the most

polymorphic V_H segments is V1-69 with 13 known alleles including duplication (Sasso *et al.*, 1993). Polymorphism may be of functional significance. One obvious possibility is expansion of repertoire. Polymorphic V_H may affect the affinity of the antibody for its ligand as even mutations in FR residues of the Ig have been shown to influence the binding affinity (Foote and Winter, 1992). Furthermore, expression of particular allelic variants could influence the efficiency of H–L chain pairing or interaction with B cell super-antigens. It is important to test whether V_H polymorphisms are associated with disease susceptibility. Some reports have suggested the association of V_H polymorphisms with autoimmune diseases such as rheumatoid arthritis, systemic lupus erythematosus and multiple sclerosis (Yang *et al.*, 1990; Walter *et al.*, 1991b), while others have reported the absence of a clear association (Hashimoto *et al.*, 1993; Shin *et al.*, 1993a).

Physical mapping of D_H and J_H segments

The human V_H locus ends with a cluster of J_H gene segments lying just upstream of the C_μ gene (Fig. 3). The human J_H cluster contains three pseudo J_H segments interspersed among six functional J_H segments (Ravetch *et al.*, 1981). A human counterpart to the murine DQ52 D_H segment exists about 100 bp 5′ of the first functional J_H (Ravetch *et al.*, 1981). A number of additional human D_H segments have been identified, including ones homologous to the murine DFL16 segments as well as a number of those that are markedly dissimilar in size and sequence (Siebenlist *et al.*, 1981; Schroeder *et al.*, 1987; Buluwela *et al.*, 1988; Ichihara *et al.*, 1988a,b; Zong *et al.*, 1988; Sonntag *et al.*, 1989; Shin *et al.*, 1993b). Initially, a family of D_H segments

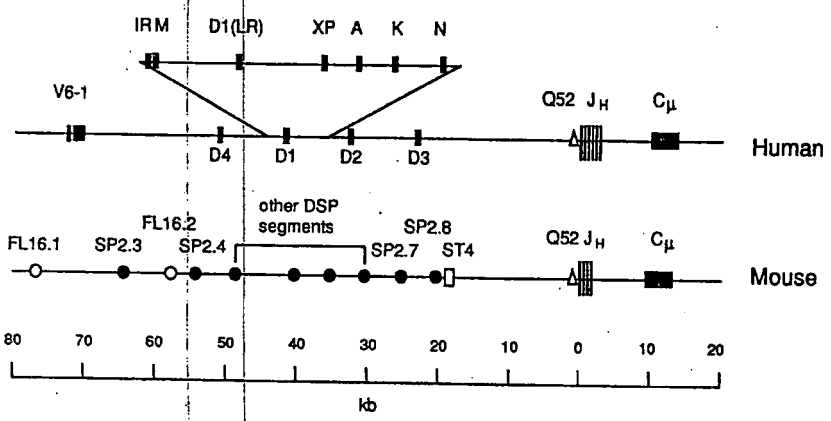


Fig. 3. Comparative map of human and mouse D–J_H loci. Mapping information of human V and D segments is from Siebenlist *et al.* (1981), Matsuda *et al.* (1988), Sato *et al.* (1988) and Ichihara *et al.* (1988a,b). Mouse part is modified from Feeney and Riblet (1993).

(D1–D4) was shown to be encoded at 9-kb regular intervals between V_H and J_H clusters (Siebenlist *et al.*, 1981). Subsequently, six novel D_H segments (IR, M, XP, A, K and N) were identified by nucleotide sequencing analysis of a 15-kb DNA fragment containing the D1 segment and its surrounding region (Ichihara *et al.*, 1988a,b) (Fig. 3). Each D_H cluster (D1–D4) appears to contain seven D_H segments (Buluwela *et al.*, 1988; Ichihara *et al.*, 1988a,b; Shin *et al.*, 1993b). Taken together, the estimated total number of D_H segments on chromosome 14 would be about 28. Frequent polymorphic deletion of the D1 segment was found in the Japanese population (Zong *et al.*, 1988). The most proximal known D_H segment of a human D_H family (D21/9) is 20 kb upstream of the DQ52/J_H cluster (Buluwela *et al.*, 1988) in a position analogous to that of the proximal DSP2 segment of the mouse. Comparison of the nucleotide sequence of *Suncus murinus*, human and mouse from a position upstream of DQ52 to the S_μ region indicated that D_H and J_H segments, consisting of coding and signal regions, are highly conserved (Okamura *et al.*, 1993). Moreover, although extensive sequence homology in the region between J_H and S_μ was observed between mouse and human, only core portions of the enhancer region of *Suncus murinus* exhibited homology to those of mouse and human. Sequence conservation of J_H segments in *Suncus murinus*, mouse and human was observed not only at the amino acid level but also at the nucleotide level including the third letters of the codons, which is difficult to explain by selection of the protein structure.

V_H and D_H segments on chromosome 15 and 16

Although the V_H locus is located at the telomere end of chromosome 14q, several V_H and two D_H clusters remained unmapped for some time. The first evidence that a D_H segment is located on chromosome 15 was obtained by *in situ* hybridization (Chung *et al.*, 1984). Subsequently, studies using *in situ* hybridization as well as human/rodent somatic hybrid cells (Cherif and Berger, 1990; Matsuda *et al.*, 1990; Nagaoka *et al.*, 1994; Tomlinson *et al.*, 1994) identified two V_H orphon loci on chromosome 15q11 and chromosome 16p11.

Studies on cosmid and YAC clones derived from these orphon loci revealed several striking findings (Matsuda *et al.*, 1990; Nagaoka *et al.*, 1994). First, about 40% of V_H segments in both loci (three out of seven V_H on chromosome 16 and one out of three V_H on chromosome 15) are apparently functional. A totally different approach based on PCR, using somatic cell hybrid DNAs as templates, specifically amplified 24 V_H segments including 10 apparently functional ones on chromosomes 15 and 16 (Tomlinson *et al.*, 1994). Second, putative origins for the orphon V_H segments on chromosomes 15 and 16 were found in the 0.43–0.25 Mb J_H-proximal V_H region on chromosome 14 (Fig. 4). Comparison of the corresponding V_H segments suggests that a DNA fragment of more than 100 kb might have been translocated simultaneously to chromosomes 15 and 16 approximately 20 million years ago. Four overlapping YAC clones covering the V_H orphon locus on chromosome 16 were isolated, and seven V_H segments were identified and sequenced. All of

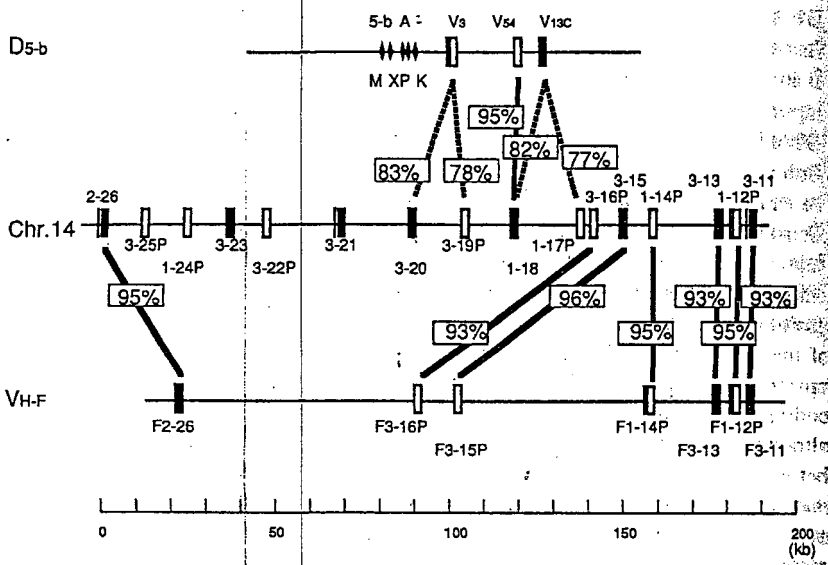


Fig. 4. Comparison of V_H segments on chromosomes 15 and 16 with their counterparts on chromosome 14. (Modified from Nagaoka *et al.*, 1994.)

seven orphon V_H segments have more than 93% identity with the corresponding V_H segments on chromosome 14. The most remarkable homology was found between two truncated pseudogenes, VF1-12P and V1-12P, in which the homology extends into the region 3′ to the truncation site. The homology between the orphon V_H segments on chromosome 15 and the corresponding V_H segments on chromosome 14 is less remarkable except for one pair (V54/V1-18). *In situ* hybridization studies using cosmid clones confirmed that two orphon loci are located on chromosome 15q11–q12 and chromosome 16p11 (Nagaoka *et al.*, 1994; Tomlinson *et al.*, 1994). The orphon locus on chromosome 15 appears to contain at least four clusters of D_H segments, each of which consists of five D_H segments (Matsuda *et al.*, 1990; Nagaoka *et al.*, 1994). One of the D_H clusters (D5-b) is flanked by three V_H segments. Interestingly, these three V_H segments are located 3′ to the D5-b cluster. The polarity of one of them (V3) (Matsuda *et al.*, 1988, 1990; Nagaoka *et al.*, 1994) was determined and shown to have the same transcriptional orientation relative to D_H.

V_H segment usage and repertoire formation

Compelling evidence indicates that V_H, D_H and J_H segments are not used equally. Biased usage of particular segments during early phases of ontogeny was first reported in mouse (Yancopoulos *et al.*, 1984; Reth *et al.*, 1986). Similarly, dominant expression

of V1-6, D_HQ52 and J_H4 (Berman *et al.*, 1991; Pascual *et al.*, 1993) in early ontogeny was demonstrated in human. One study examined the V_H segments expressed in 14 and 10 independent H-chain cDNA sequences isolated from 130-day and 104-day human fetal liver cDNA libraries, respectively (Schroeder *et al.*, 1987; Schroeder and Wang, 1990). Notably, six of these sequences employed an identical V_H segment of the V_H3 family (56P1 or V3-30), indicating that the early human repertoire is biased.

Utilization of V_H segments may be influenced by a number of factors that can be grouped into (a) those affecting the recombination frequency and (b) those affecting selection of B cells expressing that particular V_H segment. Group (a) includes distance between D_H (or J_H) and V_H segments, variation in the recombination signal sequence, and locations that favour the recombinase accessibility. Group (b) includes self-antigens and bacterial super-antigens.

The initial observation of the preferential usage of J_H-proximal V_H segments in mouse led to the hypothesis that the proximity of V_H segments to J_H favours biased expression of V_H segments in early stages of ontogeny. The complete physical mapping has clearly shown that the location of V_H segments within the locus has little association with the frequency of V_H usage in human. V_H segments often used preferentially in the early stages of ontogeny are V6-1, V1-2, V2-5, V3-13, V3-15, V3-23, V3-30, V5-51, V3-53 and V4-59 segments (Matsuda *et al.*, 1993). The V1-69 segment is also used frequently in peripheral B cells (Schwartz and Stoller, 1994), B-cell leukaemia, and autoantibodies (Zouali, 1992). This V_H segment is highly related to 51P1 cDNA, which is also found in the fetal repertoire and is localized approximately 900 kb upstream of the J_H cluster (Cook *et al.*, 1994). The results indicate that V_H segments preferentially used in early stages of ontogeny do not necessarily cluster in the J_H-proximal region.

Matsuda *et al.* (1993) have identified several germline V_H segments that were often used for autoantibodies, although it is premature to conclude that only limited V_H segments could be used for autoantibodies. The V3-30 sequence was homologous to cDNA for rheumatoid factors, RF-TS2, RF-SJ1 and RF-SJ2 (Pascual *et al.*, 1990). Also, this germline V_H segment was 99.7% identical to cDNA for Kim 1.6 autoantibody (Cairns *et al.*, 1989), which has DNA-binding activity. The V3-30 segment is also the germline counterpart of 56P1 cDNA, which is most frequently expressed in the fetal repertoire (Schroeder *et al.*, 1987; Schroeder and Wang, 1990). Similarly, the V3-15 segment, which is the germline gene of 20P1 cDNA expressed in fetal liver, is 99.7% identical to cDNA for 4B4, an anti-Sm antibody (Sanz *et al.*, 1989). The V3-23 segment is identical to 18/2 (an anti-DNA autoantibody) (Dersimonian *et al.*, 1987) and 30P1 cDNA found in fetal liver (Schroeder *et al.*, 1987; Schroeder and Wang, 1990). Such correlation between autoantibody V_H and early repertoire V_H may indicate that the preferred usage of V_H segments in early stages of ontogeny is due to positive selection by self-antigens rather than J_H-proximal location of the V_H segments.

In any case, the complete physical map of the human V_H locus has contributed to the identification of germline origins of autoantibodies. Comparison of V_H usage among polymorphic individuals may also shed light on mechanisms for biased V_H usage.

Pseudogenes and gene conversion

Are abundant, conserved pseudo V_H and orphon V_H segments of any functional significance? Since some orphon V_H segments are apparently functional, they can be joined, in theory, to J_H segments on chromosome 14 through interchromosomal recombination. As a V_H–D_H fusion product was isolated from a human B-cell line (Shin *et al.*, 1993b), a similar fusion of V_H–D_H can be formed on chromosome 15. In addition, germline transcripts of orphon V_H have been identified in human fetal liver (Cuisinier *et al.*, 1993), suggesting that orphon V_H loci might be targets of recombinase. Unfortunately, however, no direct evidence for the expression or recombination of orphon V_H segments has been reported so far. Conserved pseudogenes have been already shown to serve as sequence donors for gene conversion in other species. Somatic gene conversion (or double unequal crossing-over) has been shown to take place to amplify the V-region repertoire in chicken (Raynaud *et al.*, 1987; Tompson and Neiman, 1987) and rabbit (Becker and Knight, 1990) but not in mouse and human.

To explain the extensive conservation of V_H pseudogenes, evidence for germline gene conversion was looked for among V_H segments located in the 3′ 0.8-Mb region (Haino *et al.*, 1994). To screen the candidates of gene conversion events, the substitution rates in the intron and the synonymous position of the coding region were compared in pairs of V_H segments. This is because the introns and the synonymous positions have been shown to evolve at high and remarkably similar rates in different genes (Miyata *et al.*, 1980; Hayashida and Miyata, 1983). In addition, both introns and synonymous positions behave like clocks as they accumulate base substitutions at approximately constant rates with respect to geological time. A clear difference in substitution rates between the intron and synonymous position of a given pair of V_H segments suggests some recombination events.

Haino *et al.* (1994) selected for comparison several pairs of V_H segments that appear to be recently duplicated. DNA sequence homology of each pair (V3-62P/V3-60P and V4-61P/V4-59) of the tandemly duplicated segments (Kodaira *et al.*, 1986) is greater than 94%, which is much higher than the homology to other published V_H segments belonging to the same V_H family. V3-62P and V3-60P have the same mutation in the heptamer signal sequence and the same 3-bp deletion in the 23-bp spacer of the recombination signal, suggesting that the internal duplication event occurred after these deleterious mutations. When V3-62P and V3-60P are compared using the above evolutionary molecular clock, the nucleotide difference (0.1637 ± 0.0461) at the intron (K^c_i) was significantly greater than that (0.0821 ± 0.0339) at the synonymous position of the coding region (K^c_s), indicating that the segmental change in the intron occurred in either V3-60P or V3-62P. To look for the origin of the modified sequence, intron sequences of V3-60P, V3-62P and other V_H segments were compared. When V3-43 and V3-62P are compared, K^c_i (0.0820 ± 0.0280) was very much smaller than K^c_s (0.4165 ± 0.0761). Nucleotide sequences of the leader, intron and FR1 sequences of V3-43 and V3-62P are highly homologous whereas their 3′ halves are diverged. Such unusual homology of the 5′

half region, including the intron between V3-43 and V3-62P, is most likely explained by germline gene conversion because the 5′ half sequence of V3-62P, which must have been similar to that of the duplicated partner V3-60P, appears to be unidirectionally corrected by V3-43. This example supports the hypothesis that gene conversion contributes to the maintenance of the pseudogene structure. Since many conserved pseudogenes can be either acceptors or donors of germline gene conversion, the high percentage of germline pseudogenes in the V_H segments should also contribute to germline V_H diversity. The authors cannot completely exclude the possibility that double unequal crossing-over took place between V3-43 and V3-60P and a modified allele of V3-43 had been lost.

Evolution of the V_H loci

Almost all animals that have Ig carry multiple V_H segments, indicating that duplication of V_H segments must have started quite a long time ago. It is also important to realize that reorganization of the V_H locus is still continuing as evidenced by dramatic differences in V_H locus organization between mouse and human. The recent translocation of V_H and D_H segments to chromosomes 15 and 16 is further evidence for dynamic reshuffling of the V_H locus.

Matsumura *et al.* (1994) looked for genetic traits that may allow steps of DNA rearrangement within the human V_H loci to be traced. They have used 14 non-repetitive intergenic probes that can detect two to seven cross-hybridizing bands within the 3′ 0.8-Mb region of the V_H locus on chromosome 14. Most of these probes also detected a few bands on chromosome 15 or 16, further confirming the recent translocation of the orphon loci. Such studies identified several pairs of regions that are hybridized by an identical set of non-repetitive probes. Each of five different sets of probes hybridized as clusters to two or three regions in a dispersed manner. In most of the cases, V_H segments adjacent to homologous clusters are closely related. The dispersed appearance of these clusters of non-repetitive sequences indicated that translocation of DNA fragments frequently took place in the human V_H locus. Careful comparison of V_H sequences revealed obvious tandem duplication of sets of V_H segments: V3-33–V4-34/V3-30–V4-28 and V3-62P–V4-61/V3-60P–V4-59 (Kodaira *et al.*, 1986; Matsuda *et al.*, 1993). The longer duplication set is also associated with duplication of non-repetitive probes (Matsumura *et al.*, 1994). Among all these pairs of translocation and duplication none were inverted. Such non-repetitive probes should be useful not only to trace the evolution of the V_H locus but also to investigate polymorphisms of this locus. Extensive RFLP analyses using the probes isolated in this study would give us many markers to test the genetic linkage between RFLPs and susceptibility to immune disorders.

The human genome contains a large fraction of interspersed repetitive sequences such as short interspersed elements (*Alu* repeats) and long interspersed elements (L1 repeats). The total numbers of *Alu* and L1 repeats in the human genome have been estimated to be $3–5 \times 10^5$ and $10^4–10^5$, respectively. Assuming random distribution of

these repeats throughout the genome, it is expected that there are 70–110 *Alu* repeats and 2–20 L1 repeats within the 730-kb region analysed in this study. Because recent studies have demonstrated that some repetitive sequences could be hotspots for recombination in the genome (Hyrien *et al.*, 1987; Devlin *et al.*, 1990), the frequent reorganization of the human V_H locus may be associated with the content and distribution of these repetitive sequences.

However, comparison between the homologous (and possibly translocated) V_H segments and distribution of repetitive sequences has shown that flanking regions of closely related V_H segments do not have a similar distribution to *Alu* and L1 repeats and that these V_H surrounding regions are not necessarily abundant in repetitive sequences. These studies failed to provide evidence that the homologous recombination mediated by repetitive sequences might be the main driving force of frequent reorganization of the V_H locus. Most *Alu* repeats were reported to have been amplified within the last 60 million years (Shen *et al.*, 1991). In particular, members of the HS subfamily, a subfamily of human *Alu* repeats, have spread after the divergence of chimpanzee and human (within the last 5 million years) (Shen *et al.*, 1991). Because the closely related V_H segments were estimated to have been generated 55 to 22 million years ago (Matsumura *et al.*, 1994), the most likely explanation in this case is that many *Alu* repetitive elements transposed into random positions after V_H reorganization.

MOUSE V_H LOCUS

V_H subgroup and family

Murine V_H gene segments were originally divided into three major subgroups on the basis of protein sequences (reviewed by Kabat *et al.*, 1991). These three protein subgroups have been further subdivided to yield 14 distinct V_H families based on nucleotide sequence relatedness according to the definition described for the human V_H family (Brodeur and Riblet, 1984; Winter *et al.*, 1985; Kofler *et al.*, 1992). Human subgroups I, II and III are homologous to mouse subgroups II, I and III, respectively (Table II). Human V_H4, V_H2, V_H1 and V_H7 families correspond to mouse 3660, 3609, J558 and VGAM 3–8 families, respectively, based on 70% nucleotide sequence homology as described previously (Lee *et al.*, 1987; Berman *et al.*, 1988; Haino *et al.*, 1994). The human V_H3 family appears to correspond to most of mouse subgroup III families, including 7183, T15, J606, X24, DNA4, CP3 and 3609N. The V_H6 family is closest to the 3660 family. The human V_H5 family does not have any mouse counterparts that are more than 70% homologous. Conversely, the mouse Q52 family does not show more than 66% homology with any of the human V_H families. It is interesting to note that the 7183 family members are more than 80% homologous to human V_H3 family members except for V3-15 and V3-49. The V3-15 and V3-49 sequences are 79 and 82%, respectively, homologous to the V11 sequence of the

S107 family. In addition, almost all human V_H3 family members are more than 70% homologous to mouse subgroup III families. The results indicate that the V_H3 family is more conserved than other families between mouse and human, which could be due to some functional constraint.

Since few physical mapping data are available, the estimation of mouse V_H segments is far from accurate. Initial estimations were based on the count of restriction fragments hybridizing to a given V_H probe. Such estimation assumes that each restriction fragment of 10–20 kb may contain a single V_H and overlapping restriction fragments may be negligible. By this approach, the J558 V_H family consisted of 50–100 or more specific hybridizing fragments (Brodeur and Riblet, 1984; Schiff *et al.*, 1985) whereas other families such as S107 have only a few hybridizing fragments (Crews *et al.*, 1981; Brodeur and Riblet, 1984). These approaches have led to estimates of the size of the murine germline V_H repertoire of approximately 100 or so members (Brodeur and Riblet, 1984). A very rough estimate of members belonging to the 14 known V_H families is shown in Table II. However, this quantitative approach may give an underestimate, as suggested by observations that V_H sequences may sometimes be more closely spaced and by the existence of multiple unique V_H-containing restriction fragments of the same size (Bothwell *et al.*, 1981; Siekevitz *et al.*, 1983; Schiff *et al.*, 1985; Berman *et al.*, 1988; Rathbun *et al.*,

Table II V_H classifications and V_H gene repertoire

Mouse V_H families*	Protein subgroups†	Complexity†	Related human V_H families
Q52 (V_H2)	I	15	
3660 (V_H3)	I	5–8	V_H4, V_H6
3609 (V_H8)	I	7–10	V_H2
CH27 (V_H12)	I	1	V_H4
J558 (V_H1)	II	60–1000	V_H1
VGAM3-8 (V_H9)	II	5–7	V_H7
SM7 (V_H14)	II	3–4	V_H1
X-24 (V_H4)	III	2	
7183 (V_H5)	III	12	
J606 (V_H6)	III	10–12	
S107 (V_H7)	III	2–4	V_H3
MRL-DNA4 (V_H10)	III	2–5	
CP3 (V_H11)	III	1–6	
3609N (V_H13)	III	1	

* V_H gene families 1–7 (Brodeur and Riblet, 1984), 8 and 9 (Winter *et al.*, 1985), 10 (Kofler, 1988), 11 (Reininger *et al.*, 1988), 12 (Pennell *et al.*, 1989), 13 and 14 (Tutter *et al.*, 1991). A prototype member for each family is given with the family number in parentheses. In this study, V_H families 3609N and SM7 have been tentatively termed V_H13 and 14, respectively. The families have been organized into three phylogenetically related groups (Tutter and Riblet, 1988).

† According to Kabat *et al.* (1991).

‡ Estimated number of V_H genes per family in the germline. For references see footnote* and Livant *et al.* (1986), Dzierzak *et al.* (1986), Perlmutter *et al.* (1984), Siu *et al.* (1987) for V_H1, 3, 6 and 7, respectively.

1988). Other hybridization-based approaches have included solution hybridization experiments, which suffer from a lack of ideal kinetics when measuring hybridization to a spectrum of partially related sequences (Livant *et al.*, 1986), and plaque hybridization experiments, which determine the number of clones from a given V_H family relative to single copy sequence clones in a single-genome equivalent of a genomic DNA library (Livant *et al.*, 1986; Berman *et al.*, 1988). The latter approaches have suggested that the murine V_H locus, at least in some strains, may contain 1000 or more members, most of these being contributed by the exceptionally large J558 V_H family, which in BALB/c mice may contain 500–1000 or more members by itself.

While the size of the V_H gene repertoire is relatively conserved between different inbred strains of mice, there are still some important inter-strain differences in certain families, particularly J558 (discussed in Meek *et al.*, 1990). The exact number of members in a particular strain is known for some V_H gene families, and for most others estimates are within a relatively narrow range (Table II). Only the size of the largest family (J558) is still controversial, with estimates varying between 60 (Brodeur and Riblet, 1984) to >1000 (Livant *et al.*, 1986) members. However, since only about 30–50% of adult mitogen-stimulated splenocytes express the J558 family, the actual J558 family size might be closer to 60 than to 1000. Alternatively (or in addition), there may be multiple non-functional or essentially identical J558 family segments, which would explain the relative under-representation of this family in the expressed repertoire.

V_H segment organization

Murine V_H families are generally organized into clusters of related V_H genes. The suggestion that mouse V_H segments belonging to the same family are clustered was first made by Kemp *et al.* (1981) after they isolated distinct genomic DNA clones, each bearing pairs of highly related V_H sequences (S107 and J606). Bothwell *et al.* (1981) and Givol *et al.* (1981) reported similar findings for members of the J558 family. So-called deletion-mapping analyses of B cell lines that had rearranged known V_H segments (Kemp *et al.*, 1981; Rechavi *et al.*, 1982; Reth *et al.*, 1986; Blankenstein and Krawinkel, 1987; Rathbun *et al.*, 1987) were consistent with a clustered organization of V_H segment family members. V_H segment analyses of IgH-recombinant mouse strains also indicated a generally clustered organization of V_H families (Brodeur *et al.*, 1984). However, other studies revealed that several V_H family members are interspersed (Crews *et al.*, 1981). In particular, pairs of families mapped at both ends of the V_H locus, namely J558 and 3609 at the 5′ end and Q52 and 7183 at the 3′ end, were shown to be intermingled extensively with each other (Reth *et al.*, 1986; Blankenstein and Krawinkel, 1987; Rathbun *et al.*, 1987). Furthermore, J558-related segments appear to be dispersed over a wide range in the V_H locus.

Deletion-mapping analyses together with strain-specific RFLP analyses of V_H families in recombinant inbred strains were used to order eight V_H families and

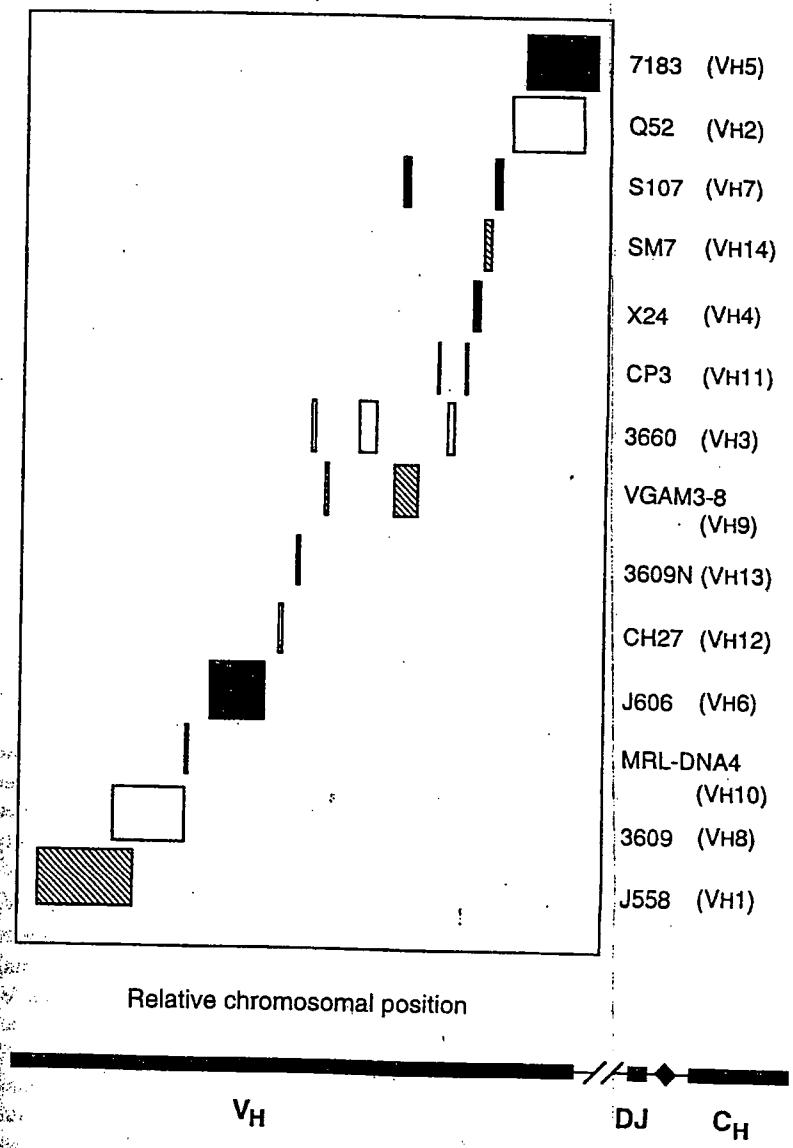


Fig 5. Relative chromosomal position of 14 murine V_H families. Mapping information of V_H10, 11, 12, 13 and 14 families is from Meek *et al.* (1990), Pennell *et al.* (1989) and Tutter *et al.* (1991). V_H families corresponding to V_H1, 2 and 3 protein subgroups are indicated by open, shaded and filled rectangles, respectively. (Modified from Brodeur *et al.*, 1988.)

generate a map of 5′–(3609, J606, 3660, X24)–J558–S107–Q52–7183–D_H–J_H–3′ (Kemp *et al.*, 1981; Rechavi *et al.*, 1982; Brodeur *et al.*, 1984; Mäkelä *et al.*, 1984; Riblet *et al.*, 1987) for the BALB/c mouse strain; the V_H families in parentheses were not mapped relative to each other. In A/J mice the nine V_H families have been positioned relative to specific rearranged V_H genes from the J558 and 3660 V_H clusters, using deletion mapping, to generate a V_H map order of 5′–3609–J558–(J606, VGAM3–8, S107)–3660–(X24, Q52, 7183)–D_H–J_H–3′ (Rathbun *et al.*, 1987); BALB/c and C57BL/6 B-cell lines that rearranged J558 V_H segments were consistent with the A/J map order. More recently, Brodeur *et al.* (1988) mapped 13 clusters of nine V_H families utilizing 32 Abelson murine leukaemia virus (A-MuLV)-transformed pre-B cell lines that had undergone VDJ recombination. In these cell lines 51 chromosomes had rearranged and provided useful information for mapping nine V_H families, which showed that the order is 5′–J558–3609–J606–3660–VGAM3.8–3660–S107–VGAM3.8–3660–X24–S107–Q52–7183–D_H–J_H–3′. These authors have also shown that several V_H segments are split into few clusters as shown in Fig. 5. This map is slightly different from their previous mapping and more similar to the A/J map. The difference could be explained by confusion due to the partially dispersed distribution of large families like J558. A larger number of rearranged chromosomes examined in the study by Brodeur *et al.* (1988) lend weight to their conclusion. Strain polymorphism could be another source of difference. However, as will be discussed below, the general organization of the V_H locus appears to be more conserved than previously expected among haplotypes.

Subsequently, five more new families were identified and mapped: V_H10 (Kofler, 1988; Meek *et al.*, 1990), V_H11 (Reininger *et al.*, 1988; Hardy *et al.*, 1989; Meek *et al.*, 1990), V_H12 (Pennell *et al.*, 1989), V_HSM7 (Tutter *et al.*, 1991) and V_H3609N (Tutter *et al.*, 1991). A summary of these results is shown in Fig. 5. Since Brodeur *et al.* (1988) utilized F_1 mice carrying the Igh^a and Igh^b parent haplotypes, they could simultaneously map BALB/c and C57/BL haplotypes. These data indicate that the general organization of Igh^a and Igh^b haplotypes are almost identical. More limited data for the IgH^a, IgH^c, IgH^j haplotypes are also consistent with the V_H family positions shown above.

Unfortunately, the resolution of the technique utilized for studying mouse V_H segment organization is limited. The details of murine V_H segment organization should be elucidated by PFG mapping and direct-linkage studies using overlapping cosmid, phage or YAC clones, as has been done more extensively for human V_H families.

D_H and J_H segments

The heavy-chain V-gene segments (V_H, D_H and J_H) are arranged in contiguous but separate clusters on chromosome 12 in the mouse genome, although there is no direct physical linkage between V_H and D_H segments (Fig. 3). Four J_H segments lie at several hundred base pair intervals approximately 7 kb upstream of the C_μ gene. The

D_H segments have been subdivided into three families on the basis of coding and flanking region relatedness (Kurosawa and Tonegawa, 1982). A single DQ52 segment resides about 750 bp 5′ of J_H1; nine DSP2 segments are positioned 10–80 kb upstream of DQ52 (Wood and Tonegawa, 1983). The most 5′ DSP2 (DSP2.3) segment is positioned between the two DFL16 segments; DFL16.1 (in BALB/c) has been identified as the most upstream D_H segment characterized to date. Recently, a new functional D_H segment (DST4), which is not related to any of the known D_H families, was identified and mapped between the 3′-end DSP2 (DSP2.8) and DQ52 segments (Feeney and Riblet, 1993). All known murine V_H segments are located upstream of DFL16.1 and to date there is no evidence for additional D_H segments further upstream (within the V_H locus) or elsewhere in the genome (Ichihara *et al.*, 1989; Feeney and Riblet, 1993).

V_H segment usage and repertoire formation

The first evidence that the J_H-proximal V_H (7183) is preferentially used came from studies on nine A-MuLV-transformed pre-B lines that were supposed to be generated without selection of surface Ig as pre-B cells do not produce L chains and no surface Ig. Yancopoulos *et al.* (1984) examined nine A-MuLV-transformed BALB/c fetal liver-derived pre-B cell lines and documented 12 $V_HD_HJ_H$ rearrangements in these cell lines, of which 11 involved the 7183 family. Similar results were obtained by Perlmutter *et al.* (1985), who found that seven of nine fetal liver hybridomas express the 7183 family. In contrast to initial reports, several groups (Reth *et al.*, 1986; Lawler *et al.*, 1987; Sugiyama *et al.*, 1987; Osman *et al.*, 1988) have observed the preferential utilization of both 7183 and Q52 family members in pre-B cells. In particular analyses of A-MuLV-transformed pre-B lines that actively continued V_H to D_HJ_H joining in culture provided a useful system for the study of V_H utilization in the absence of selection by the environment. Such lines from several different mouse strains all showed frequent utilization of the J_H-proximal V_H (Reth *et al.*, 1986). Taken together, the findings (Reth *et al.*, 1986; Rathbun *et al.*, 1987) that 7183 and Q52 family members are interspersed in a variety of haplotypes, and that both families are preferentially utilized in pre-B cells, support the suggestion of Yancopoulos *et al.* (1984) of a position-dependent rearrangement of V_H segments in early ontogeny. This view, however, is not generally accepted and a locus outside of the V_H region has been suggested as being responsible for V_H gene family utilization (Wu and Paige, 1988; Atkinson *et al.*, 1991). Unfortunately, no parallel results were obtained in human V_H usage except that the J_H-proximal V_H (i.e. V6-1) is preferentially used in early ontogeny (Berman *et al.*, 1991). The difference will be resolved only when the precise physical map of the murine V_H locus is available.

In contrast to V_H segments, V_κ segment rearrangement in the newly generated repertoire appears position independent, although not entirely stochastic (Lawler *et al.*, 1989). In adult bone marrow-derived pre-B cell lines, a strong bias for rearrangement of genes from the V_κ4/5 family has been reported (Kalled and Brodeur, 1990).

Thus, primary V-gene selection might be biased in differentiating pre-B cells at all stages of ontogeny, although the as yet unidentified molecular mechanisms acting on the V_H and V_κ loci might be distinct.

In contrast to the results from pre-B cell lines, hybridomas or B-cell colonies derived from bacterial lipopolysaccharide (LPS)-activated spleen cells of adult mice had 'random' utilization of V_H families; thus, family representation depended on the family complexity (Dildrop *et al.*, 1985; Schulze and Kelsoe, 1987). Assays of V_H utilization patterns in RNA prepared from organs of normal B-cell differentiation (neonatal liver versus adult spleen) indicate contrasting V_H usage profiles (non-random versus random) in agreement with studies using cell lines. Several other groups have also reported that V_H family utilization is normalized in the adult peripheral repertoire and roughly corresponds to the germline complexity of these families (Dildrop *et al.*, 1985; Schulze and Kelsoe, 1987; Jeong and Teale, 1988; Yancopoulos *et al.*, 1988). The issue is, however, still controversial: non-random V_H gene expression in the peripheral repertoire was shown in at least one report (Lawler *et al.*, 1987), and suggested on the basis of inter-strain repertoire comparisons in others (Jeong *et al.*, 1988; Sheehan and Brodeur, 1989). Moreover, Wu and Paige (1986) did not find significant differences between newly generated and functional peripheral repertoires.

C_H LOCUS

Structure of C_H genes

All the human and murine C_H genes have been isolated and sequenced completely; the references for the complete C_H gene sequences are summarized in Table III. The mouse C_H genes for secretory forms are composed of three (α) or four (μ, δ, γ and ε) exons, each encoding a functional and structural unit of the H chain, namely a domain (Edelman *et al.*, 1969) or hinge region. In addition, one (α) or two (other) separate exons encode the hydrophobic transmembrane and short intracytoplasmic segments that are used for a membrane-form Ig. The C_α genes are exceptional because the hinge region is encoded by the C_H2 exon. The other exception is the C_δ gene, which has two additional exons, the most 3′ of which encodes a C-terminal tail for the secretory form 1 kb 5′ to the membrane exons. The size of each C_H exon is similar to that of the C_L exon, suggesting that the C_H gene evolved by duplication of a primordial single exon gene like the C_L gene. Such exon–intron organization of the C_H gene is consistent with the domain hypothesis that the H-chain protein consists of a tandem array of three or four functional units (Edelman *et al.*, 1969). The total length of each C_H gene is therefore variable, ranging from 3 to 7 kb.

Expression of the membrane exons is controlled by differential splicing. Transcripts of the membrane exons are spliced to the domain exons nearest the 3′ end

Table III References for complete nucleotide sequences of C_H genes*

Genes	Reference
Mouse	
μ	Kawakami *et al.* (1980); Rogers *et al.* (1980)
δ	Tucker *et al.* (1980); Cheng *et al.* (1982)
γ1	Honjo *et al.* (1979); Tyler *et al.* (1982)
γ2a	Yamawaki-Kataoka *et al.* (1981, 1982)
γ2b	Yamawaki-Kataoka *et al.* (1980, 1982)
γ3	Wels *et al.* (1984); Komaromy *et al.* (1983)
ε	Ishida *et al.* (1982); Liu *et al.* (1982)
α	Tucker *et al.* (1981); Word *et al.* (1983)
Human	
δ	Milstein *et al.* (1984); White *et al.* (1985)
γ1	Ellison *et al.* (1982)
γ2	Ellison and Hood (1982)
γ3	Huck *et al.* (1986)
γ4	Ellison *et al.* (1981)
ψγ	Bensmana *et al.* (1988)
ε3, ε2, ε3	Max *et al.* (1982); Ueda *et al.* (1982)
α1, α2	Flanagan *et al.* (1984)

* Corrected sequences are found in Kabat et al. (1991).

by removing the last few residues of the secreted Ig tail. All the membrane segments except C_α are encoded by two exons. The hydrophobic transmembrane segment of 26 residues is relatively conserved among all the H chains, suggesting the possibility that membrane-form Ig is anchored by a common membrane protein (Yamawaki-Kataoka *et al.*, 1982). Since the intracytoplasmic segments of the membrane-form Ig are too short (27 residues for C_γ and C_ε chains, 14 residues for C_α and two residues for C_μ and C_δ) to catalyse any enzymic activity such as phosphorylation, transduction of the triggering signal of the antigen–antibody interaction may require involvement of at least one other protein. This hypothesis has been verified by subsequent identification of Igα and Igβ proteins (see Chapter 6).

Allotypes of Ig are mostly explained by polymorphism in C_H regions. Allotypes were originally defined by antigenic differences in Ig between different strains of mice and it is not always clear which difference in the amino acid sequence is responsible for a particular allotype defined by an antibody. The typical cases are the comparison of a and b haplotypes of the $C_\gamma2a$ and $C_\gamma2b$ genes, which revealed differences in 54 and 4 residues, respectively (Ollo and Rougeon, 1982, 1983). Further studies using *in vitro* mutagenesis and expression of mutants in culture cells will define polymorphic differences responsible for antigenicity for various antibodies. The Am determinants in human and allotypic determinants in rabbit IgG were assigned to one or a few residue differences (Flanagan *et al.*, 1984; Martens *et al.*, 1984).

Organization of C_H genes

The mouse C_H gene locus, which is mapped to chromosome 12 (D'eustachio *et al.*, 1980), consists of eight genes that cluster in a 200 kb region (Shimizu *et al.*, 1982a). The order of the mouse C_H genes, 5′–J_H–(6.5 kb)–C_μ–(4.5 kb)–C_δ–(55 kb)–C_γ3–(34 kb)–C_γ1–(21 kb)–C_γ2b–(15 kb)–C_γ2a–(14 kb)–C_δ–(12 kb)–C_α–3′ (Shimizu *et al.*, 1982a), is consistent with the order proposed by the deletion profile of C_H genes in myelomas producing different Ig classes (Honjo and Kataoka, 1978). Unlike the C_λ gene (Miller *et al.*, 1982), the C_H genes share one set of J_H segments (Shimizu *et al.*, 1982a), which allows them to retain the same V_H gene during class switching. The mouse genome does not contain any well-conserved pseudogene of the C_H genes.

The general organization of the C_H gene locus is similar among laboratory strain mice, though there are many polymorphic differences. Some wild mice, however, have duplicated C_γ2a genes (Shimizu *et al.*, 1982b; Fukui *et al.*, 1984). Of 31 Japanese and Chinese wild mice screened 17 had similar duplication, suggesting that this duplication took place relatively recently.

The human C_H gene family is mapped to the q32 band of chromosome 14 (Kirsch *et al.*, 1982) and consists of nine functional genes and two pseudogenes. The organization of the human C_H gene cluster is different from that of mouse in that the C_γ–C_γ–C_ε–C_α unit is duplicated downstream of the C_μ–C_δ genes. In addition, a pseudo C_γ gene has been genetically mapped between the duplication unit (Bech-Hansen *et al.*, 1983). The 5′ C_ε or C_ε2 gene is a truncated pseudogene. The other pseudogene C_ε3 is processed and translocated to chromosome 9 (Battey *et al.*, 1982). The organization of the human C_H locus is as follows: 5′–J_H–(8 kb)–C_μ–(5 kb)–C_δ–(~60 kb)–C_γ3–(26 kb)–C_γ1–(19 kb)–Cε2–(13 kb)–C_α1–(~35 kb)–ψC_γ–(~45 kb)–C_γ2–(18 kb)–C_γ4–(23 kb)–C_ε1–(10 kb)–C_α2–3′ (Ravetch *et al.*, 1981; Flanagan and Rabbitts, 1982; Bottaro *et al.*, 1989; Hofker *et al.*, 1989).

Several deletion mutations in the C_H gene locus have been reported; deleted regions are C_γ1–C_γ4 (Lefranc *et al.*, 1982), C_γ2–C_γ4 (Migone *et al.*, 1984) or C_α1–C_ε1 (Migone *et al.*, 1984). It is rather surprising that individuals with deletions of several C_H genes have not shown any severe clinical symptoms, suggesting that C_γ and C_α subclass genes are capable of substituting for each other and that the C_γ genes might not be obligatory but might facilitate efficient protection from parasite infection.

Other mammals

Rat C_H gene organization was strikingly homologous to that of mouse. The order of the C_H genes is 5′–J_H–C_μ–C_δ–(C_γ2c/C_γ2a)–C_γ1–C_γ2b–C_ε–C_α–3′. However, rat C_γ2a and C_γ1 genes, which are very similar to each other, are most homologous to the mouse C_γ1 gene. Rat C_γ2b gene is most homologous to mouse C_γ2b and C_γ2a genes (Brüggemann *et al.*, 1986). These results suggest that although C_H loci of these species have the same number of C_γ genes, these C_γ genes have evolved very dynamically by duplication (and deletion) as shown in wild mouse (Shimizu *et al.*, 1982b; Fukui *et al.*, 1984). The organization of rabbit Ig genes is described in detail in Chapter 13.

Acknowledgements

We are grateful to Ms Y. Nakamura for her assistance in preparation of the manuscript. This work was supported in part by grants from Creative Basic Research (Human Genome Program) of the Ministry of Education, Science and Culture of Japan and from the Science and Technology Agency of Japan. This work was done while TH was a Fogarty scholar-in-residence at National Institute of Health, USA.

REFERENCES

Atkinson, M.J., Michnick, D.A., Paige, C.J. and Wu, G.E. (1991). *J. Immunol.* **146**, 2805–2812.
Ballard, D.W. and Bothwell, A. (1986). *Proc. Natl Acad. Sci. USA* **83**, 9626–9630.
Battey, J., Max, E.E., McBride, W.O., Swan, D. and Leder, P. (1982). *Proc. Natl Acad. Sci. USA* **79**, 5956–5960.
Bech-Hansen, N.T., Linsley, P.S. and Cox, D.W. (1983). *Proc. Natl Acad. Sci. USA* **80**, 6952–6956.
Becker, R.S. and Knight, K.L. (1990). *Cell* **63**, 987–997.
Bensmana, M., Huck, S., Lefranc, G. and Lefranc, M.-P. (1988). *Nucleic Acids Res.* **16**, 3108.
Berman, J.E., Mellis, S.J., Pollock, R., Smith, C.L., Suh, H., Heinke, B., Kowal, C., Surti, U., Chess, L., Cantor, C.R. and Alt, F.W. (1988). *EMBO J.* **7**, 727–738.
Berman, J.E., Nickerson, K.G., Pollock, R.R., Barth, J.E., Schuurman, R.K.B., Knowles, D.M., Chess, L. and Alt, F.W. (1991). *Eur. J. Immunol.* **21**, 1311–1314.
Blankenstein, T. and Krawinkel, U. (1987). *Eur. J. Immunol.* **17**, 1351–1357.
Bothwell, A.L.M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. and Baltimore, D. (1981). *Cell* **24**, 625–637.
Bottaro, A., de Marchi, M., Migone, N. and Carbonara, A.O. (1989). *Genomics* **4**, 505–508.
Brodeur, P.H. and Riblet, R. (1984). *Eur. J. Immunol.* **14**, 922–930.
Brodeur, P.H., Tompson, M.A. and Riblet, R. (1984). In 'Regulation of the Immune System' (H. Cantor, L. Chess and E. Sercarz, eds), pp. 445–453. Alan R. Liss, New York.
Brodeur, P.H., Osman, G.E., Mackle, J.J. and Lalor, T.M. (1988). *J. Exp. Med.* **168**, 2261–2278.
Brüggemann, M., Free, J., Diamond, A., Howard, J., Cobbold, S. and Waldmann, H. (1986). *Proc. Natl Acad. Sci. USA* **83**, 6075–6079.
Buluwela, L. and Rabbitts, T.H. (1988). *Eur. J. Immunol.* **18**, 1843–1845.
Buluwela, L., Albertson, D.G., Sherrington, P., Rabbitts, P.H., Spurr, N. and Rabbitts, T.H. (1988). *EMBO J.* **7**, 2003–2010.
Cairns, E., Kwong, P.C., Misener, V., Ip, P., Bell, D.A. and Siminovitch, K.A. (1989). *J. Immunol.* **143**, 685–691.
Cheng, H.-L., Blattner, F.R., Flitzmaurice, L., Mushinski, J.F. and Tucker, P.W. (1982). *Nature* **296**, 410–415.
Cherif, D. and Berger, R. (1990). *Genes, Chromosomes and Cancer* **2**, 103–108.
Chung, J.H., Siebenlist, U., Morton, C.C. and Leder, P. (1984). *Fed. Proc.* **43**, 1486.
Cook, G., Tomlinson, I.M., Walter, G., Riethman, H., Carter, N.P., Buluwela, L., Winter, G. and Rabbitts, T.H. (1994). *Nature Genet.* **7**, 162–168.
Crews, S., Griffin, J., Huang, H., Calame, K. and Hood, L. (1981). *Cell* **25**, 59–66.
Croce, C.M., Sander, M., Martins, J., Cicurel, L., D'Ancona, G.G., Dolby, T.W. and Koprowski, H. (1979). *Proc. Natl Acad. Sci. USA* **76**, 3416–3419.
Cuisinier, A.M., Gauthier, L., Boubli, L., Fougereau, M. and Tonnelle, C. (1993). *Eur. J. Immunol.* **23**, 110–118.

Dersimonian, H., Schwartz, R.S., Barrett, K.J. and Stollar, B.D. (1987). *J. Immunol.* **139**, 2496–2501.
D'eustachio, P., Pravtceva, D., Marcu, K. and Raddle, F.H. (1980). *J. Exp. Med.* **151**, 1545–1550.
Devlin, R.H., Deeb, S., Brunzell, J. and Hayden, M.R. (1990). *Am. J. Hum. Genet.* **46**, 112–119.
Dildrop, R., Krawinkel, U., Winter, E. and Rajewsky, K. (1985). *Eur. J. Immunol.* **15**, 1154–1156.
Dzierzak, E.A., Janeway, C.A. Jr, Richard, N. and Bothwell, A. (1986). *J. Immunol.* **136**, 1864–1870.
Eaton, S. and Calame, K. (1987). *Proc. Natl Acad. Sci. USA* **84**, 7634–7638.
Edelman, G.M., Cunningham, B.A., Gall, W.E., Gottliep, P.D., Rutishauser, U. and Waxdal, M.J. (1969). *Proc. Natl Acad. Sci. USA* **63**, 78–85.
Ellison, J. and Hood, L. (1982). *Proc. Natl Acad. Sci. USA* **79**, 1984–1988.
Ellison, J., Buxbaum, J. and Hood, L. (1981). *DNA* **1**, 11–18.
Ellison, J., Berson, B.J. and Hood, L.E. (1982). *Nucleic Acids Res.* **10**, 4071–4079.
Feeney, A.J. and Riblet, R. (1993). *Immunogenetics* **37**, 217–221.
Flanagan, J.G. and Rabbitts, T.H. (1982). *Nature* **300**, 709–713.
Flanagan, J.G., Lefranc, M.-P. and Rabbitts, T.H. (1984). *Cell* **36**, 681–688.
Foote, J. and Winter, G. (1992). *J. Mol. Biol.* **224**, 487–499.
Fukui, K., Hamaguchi, Y., Shimizu, A., Nakai, S., Moriwaki, K., Wang, C.H. and Honjo, T. (1984). *J. Mol. Cell. Immunol.* **1**, 321–330.
Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. and Cohen, J.B. (1981). *Nature* **292**, 426–430.
Green, L.L., Hardy, M.C., Maynard-Currie, C.E., Tsuda, H., Louie, D.M., Mendez, M.J., Abderrahim, H., Noguchi, M., Smith, D.H., Zeng, Y., David, N.E., Sasai, H., Garza, D., Brenner, D.G., Hales, J.F., McGuiness, R.P., Capon, D.J., Klapholz, S. and Jakobovits, A. (1994) *Nature Genet.* **7**, 13–21.
Haino, M., Hayashida, H., Miyata, T., Shin, E.K., Matsuda, F., Nagaoka, H., Matsumura, R., Taka-ishi, S., Fukita, Y., Fujikura, J. and Honjo, T. (1994). *J. Biol. Chem.* **269**, 2619–2626.
Hardy, R.R., Carmack, C.E., Shinton, S.A., Riblet, R.J. and Hayakawa, K. (1989). *J. Immunol.* **142**, 3643–3651.
Hashimoto, L.L., Walter, M.A., Cox, D.W. and Ebers, G.C. (1993). *J. Neuroimmunol.* **44**, 77–83.
Hayashida, H. and Miyata, T. (1983). *Proc. Natl Acad. Sci. USA* **80**, 2671–2675.
Hofker, M.H., Walter, M.A. and Cox, D.W. (1989). *Proc. Natl Acad. Sci. USA* **86**, 5567–5571.
Honjo, T. and Kataoka, T. (1978). *Proc. Natl Acad. Sci. USA* **75**, 2140–2144.
Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Takahashi, N. and Mano, Y. (1979). *Cell* **18**, 559–568.
Huck, S., Crawford, D.H., Lefranc, M.-P. and Lefranc, G. (1986). *Nucleic Acids Res.* **14**, 1779–1789.
Hyrien, O., Debatisse, M., Buttin, G. and Robert de Saint Vincent, B. (1987). *EMBO J.* **6**, 2401–2408.
Ichihara, Y., Abe, M., Yasui, H., Matsuoka, H. and Kurosawa, Y. (1988a). *Eur. J. Immunol.* **18**, 649–652.
Ichihara, Y., Matsuoka, H. and Kurosawa, Y. (1988b). *EMBO J.* **7**, 4141–4150.
Ichihara, Y. Hayashida, H., Miyazawa, S. and Kurosawa, Y. (1989). *Eur. J. Immunol.* **19**, 1849–1854.
Ishida, N., Ueda, S., Hayashida, H., Miyata, T. and Honjo, T. (1982). *EMBO J.* **1**, 1117–1123.
Jeong, H.D. and Teale, J.M. (1988). *J. Exp. Med.* **168**, 589–603.
Jeong, H.D., Komisar, J.L., Kraig, E. and Teale, J.M. (1988). *J. Immunol.* **140**, 2436–2441.
Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S. and Foeller, C. (1991). 'Sequences of Proteins of Immunological Interest', 5th edn. NIH publications, Washington DC.
Kalled, S.L. and Brodeur, P.H. (1990). *J. Exp. Med.* **172**, 559–566.
Kawakami, T., Takahashi, N. and Honjo, T. (1980). *Nucleic Acids Res.* **8**, 3933–3945.
Kemler, I., Schreiber, E., Müller, M.M., Matthias, P. and Schaffner, W. (1989). *EMBO J.* **8**, 2001–2008.
Kemp, D.J., Tyler, B., Bernard, O., Gough, N., Gerondakis, S., Adams, J.M. and Cory, S. (1981). *J. Mol. Appl. Genet.* **1**, 245–261.
Kirkham, P.M. and Schroeder, H.W. Jr (1994). *Semin. Immunol.* **6**, 347–360.
Kirsch, I.R., Morton, C.C., Nakahara, K. and Leder, P. (1982). *Science* **216**, 301–303.
Kodaira, M., Kinashi, T., Umemura, I., Matsuda, F., Noma, T., Ono, Y. and Honjo, T. (1986). *J. Mol. Biol.* **190**, 529–541.
Kofler, R. (1988). *J. Immunol.* **140**, 4031–4034.

Kofler, R., Geley, S., Kofler, H. and Helmberg, A. (1992). *Immunol. Rev.* **128**, 5–21.
Komaromy, M., Clayton, L., Rogers, S., Robertson, S., Kettman, J. and Wall, R. (1983). *Nucleic Acids Res.* **11**, 6775–6785.
Kurosawa, Y. and Tonegawa, S. (1982). *J. Exp. Med.* **155**, 201–218.
Lawler, A.M., Lin, P.S. and Gearhart, P.J. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 2454–2458.
Lawler, A.M., Kearney, J.F., Kuehl, M. and Gearhart, P.J. (1989). *Proc. Natl Acad. Sci. USA* **86**, 6744–6748.
Lee, K.H., Matsuda, F., Kinashi, T., Kodaira, M. and Honjo, T. (1987). *J. Mol. Biol.* **195**, 761–768.
Lefranc, M.-P., Lefranc, G. and Rabbitts, T.H. (1982). *Nature* **300**, 760–762.
Liu, C.-P., Tucker, P.W., Mushinski, G.F. and Blattner, F.R. (1980). *Science* **209**, 1348–1353.
Liu, F.-T., Albrand, K., Sutchliffe, J.G. and Katz, D.H. (1982). *Proc. Natl Acad. Sci. USA* **79**, 7852–7856.
Livant, D., Blatt, C. and Hood, L. (1986). *Cell* **47**, 461–470.
Mäkelä, O., Seppala, I.J.T., Pekonen, J., Kaartinen, M., Cazenave, P.A. and Gefter, M.L. (1984). *Ann. Immunol. (Paris)* **135C**, 169–173.
Martens, C.L., Currier, S.J. and Knight, K.L. (1984). *J. Immunol.* **133**, 1022–1027.
Matsuda, F., Lee, K.H., Nakai, S., Sato, T., Kodaira, M., Zong, S.Q., Ohno, H., Fukuhara, S. and Honjo, T. (1988). *EMBO J.* **7**, 1047–1051.
Matsuda, F., Shin, E.K., Hirabayashi, Y., Nagaoka, H., Yoshida, M.C., Zong, S.Q. and Honjo, T. (1990). *EMBO J.* **9**, 2501–2506.
Matsuda, F., Shin, E.K., Nagaoka, H., Matsumura, R., Haino, M., Fukita, Y., Taka-ishi, S., Imai, T., Riley, J.H., Anand, R., Soeda, E. and Honjo, T. (1993). *Nature Genet.* **3**, 88–94.
Matsumura, R., Matsuda, F., Nagaoka, H., Shin, E.K., Fukita, Y., Haino, M., Fujikura, J. and Honjo, T. (1994). *J. Immunol.* **152**, 660–666.
Max, E.E., Battey, J., Mey, R., Kirsch, I.R. and Leder, P. (1982). *Cell*, **29**, 691–699.
Meek, K., Rathbun, G., Reininger, L., Jaton, J.C., Kofler, R., Tucker, P.W. and Capra, J.D. (1990). *Mol. Immunol.* **27**, 1073–1081.
Migone, N., Oliviero, S., de Lange, G., Delacroix, D.L., Boschis, D., Altruda, F., Silengo, L., DeMarchi, M. and Carbonara, A.O. (1984). *Proc. Natl Acad. Sci. USA* **81**, 5811–5815.
Miller, J., Selsing, E. and Storb, U. (1982). *Nature* **295**, 428–430.
Milstein, C.P., Deverson, E.V. and Rabbitts, T.H. (1984). *Nucleic Acids Res.* **12**, 6523–6535.
Miyata, T., Yasunaga, T. and Nishida, T. (1980). *Proc. Natl Acad. Sci. USA* **77**, 7328–7332.
Nagaoka, H., Ozawa, K., Matsuda, F., Hayashida, H., Matsumura, R., Haino, M., Shin, E.K., Fukita, Y., Imai, T., Anand, R., Yokoyama, K., Eki, T., Soeda, E. and Honjo, T. (1994). *Genomics* **22**, 189–197.
Okamura, K, Ishiguro, H., Ichihara, Y. and Kurosawa, Y. (1993). *Mol. Immunol.* **30**, 461–467.
Ollo, R. and Rougeon, F. (1982). *Nature* **296**, 761–763.
Ollo, R. and Rougeon, F. (1983). *Cell* **32**, 515–523.
Osman, G.E., Wortis, H.H. and Brodeur, P.H. (1988). *J. Exp. Med.* **168**, 2023–2030.
Pascual, V., Randen, I., Tompson, K., Sioud, M., Forre, O., Natvig, J. and Capra, J.D. (1990). *J. Clin. Invest.* **86**, 1320–1328.
Pascual, V., Verkruyse, L., Casey, M.L. and Capra, J.D. (1993). *J. Immunol.* **151**, 4164–4172.
Pennell, C.A., Sheehan, K.M., Brodeur, P.H. and Clarke, S.H. (1989). *Eur. J. Immunol.* **19**, 2115–2121.
Perlmutter, R.M., Klotz, J.L., Bond, M.W., Nahm, M., Davie, J.M. and Hood, L. (1984). *J. Exp. Med.* **159**, 179–192.
Perlmutter, R.M., Kearney, J.F., Chang, S.P. and Hood, L. (1985). *Science* **227**, 1597–1601.
Rathbun, G.A., Capra, J.D. and Tucker, P.W. (1987). *EMBO J.* **6**, 2931–2937.
Rathbun, G.A., Otani, F., Milner, E.C.B., Capra, J.D. and Tucker, P.W. (1988). *J. Mol. Biol.* **202**, 383–395.
Ravetch, J.V., Siebenlist, U., Korsmeyer, S., Waldmann, T. and Leder, P. (1981). *Cell* **27**, 583–591.
Raynaud, C.-A., Anques, V., Grimal, H. and Weill, J.C. (1987). *Cell* **48**, 379–388.
Rechavi, G., Bientz, B., Ram, D., Ben-Neriah, Y., Cohen, J.B., Zakut, R. and Givol., D. (1982). *Proc. Natl Acad. Sci. USA* **79**, 4405–4409.
Reininger, L., Kaushik, A., Izui, S. and Jaton, J.C. (1988). *Eur. J. Immunol.* **18**, 1521–1526.
Reth, M., Jackson, N. and Alt, F.W. (1986). *EMBO J.* **5**, 2131–2138.

Riblet, R., Brodeur, P., Tutter, A. and Tompson, M.A. (1987). In 'Evolution and Vertebrate Immunity: The Antigen Receptor and MHC Gene Families' (G. Kelsoe and D.H. Schultze, eds), p. 53. University of Texas, Austin.
Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. and Wall, R. (1980). *Cell* **20**, 303–312.
Sanz, I., Dang, H., Takei, M., Talal, N. and Capra, J.D. (1989). *J. Immunol.* **142**, 883–887.
Sasso, E.H., Willems van Dijk, K., Bull, A.P. and Milner, E.C.B. (1993). *Proc. Natl Acad. Sci. USA* **89**, 10430–10434.
Sato, T., Matsuda, F., Lee, K.H., Shin, E.K. and Honjo, T. (1988). *Biochem. Biophys. Res. Commun.* **154**, 265–271.
Schiff, C., Milili, M. and Fougereau, M. (1985). *EMBO J.* **2**, 1225–1330.
Schroeder, H.W., Jr and Wang, J.Y. (1990). *Proc. Natl Acad. Sci. USA* **87**, 6146–6150.
Schroeder, H.W., Jr, Hillson, J.L. and Perlmutter, R.M. (1987). *Science* **238**, 791–793.
Schroeder, H.W., Jr, Walter, M.A., Hofker, M.H., Ebens, A., Willem van Dijk, K., Liao, L.C., Cox, D.W., Milner, E.C.B. and Perlmutter, R.M. (1988). *Proc. Natl Acad. Sci. USA* **85**, 8196–8200.
Schroeder, H.W., Jr, Hillson, J.L. and Perlmutter, R.M. (1990). *Int. Immunol.* **20**, 41–50.
Schulze, D.H. and Kelsoe, G. (1987). *J. Exp. Med.* **166**, 163–172.
Schwartz, R.S. and Stoller, B.D. (1994). *Immunol. Today* **15**, 27–32.
Sheehan, K.M. and Brodeur, P.H. (1989). *EMBO J.* **8**, 2313–2320.
Shen, A., Humphries, C., Tucker, P. and Blattner, F. (1987). *Proc. Natl Acad. Sci. USA* **84**, 8563–8567.
Shen, M.R., Batzer, M.A. and Deininger, P.L. (1991). *J. Mol. Evol.* **33**, 311–320.
Shimizu, A., Takahashi, N., Yaoita, Y. and Honjo, T. (1982a). *Cell* **28**, 499–506.
Shimizu, A., Hamaguchi, Y., Yaoita, Y., Moriwaki, K., Kondo, S. and Honjo, T. (1982b). *Nature* **298**, 82–84.
Shin, E.K., Matsuda, F., Nagaoka, H., Fukita, Y., Imai, T., Yokoyama, K., Soeda, E. and Honjo, T. (1991). *EMBO J.* **10**, 3641–3645.
Shin, E.K., Matsuda, F., Ozaki, S., Kumagai, S., Olerup, O., Strom, H., Melchars, I. and Honjo, T. (1993a). *Immunogenetics* **38**, 304–306.
Shin, E.K., Matsuda, F., Fujikura, J., Akamizu, T., Sugawa, H., Mori, T. and Honjo, T. (1993b). *Eur. J. Immunol.* **23**, 2365–2367.
Siebenlist, U., Ravetch, J.V., Korsmeyer, S., Waldmann, T. and Leder, P. (1981). *Nature* **294**, 631–635.
Siekevitz, M., Huang, S.Y. and Gefter, M.L. (1983). *Eur. J. Immunol.* **13**, 123–132.
Siu, G., Springer, E.A., Huang, H.V., Hood, L.E. and Crews, S.T. (1987). *J. Immunol.* **138**, 4466–4471.
Sonntag, D., Weingartner, B. and Grutzmann, R. (1989). *Nucleic Acids Res.* **17**, 1267.
Sugiyama, H., Maeda, T., Tani, Y., Miyake, S., Oka, Y., Komori, T., Ogawa, H., Soma, T., Minami, Y., Sakato, N. and Kishimoto, S. (1987). *J. Exp. Med.* **166**, 607–612.
Taylor, L.D., Carmack, C.E., Huszar, D., Higgins, K.M., Mashayekh, R., Sequar, G., Schramm, S.R., Kuo, C.C., O'Donnell, S.L., Kay, R.M., Woodhouse, C.S. and Lonberg, N. (1994). *Int. Immunol.* **6**, 579–591.
Tomlinson, I.M., Walter, G., Marks, J.D., Llewelyn, M.B. and Winter, G. (1992). *J. Mol. Biol.* **227**, 776–798.
Tomlinson, I.M., Cook, G.P., Carter, N.P., Elaswarapu, R., Smith, S., Walter, G., Buluwela., L., Rabbitts, T.H. and Winter, G. (1994). *Hum. Mol. Genet.* **3**, 856–860.
Tompson, C.B. and Neiman, P.E. (1987). *Cell* **48**, 369–378.
Tucker, P.W., Liu, C.P., Mushinski, J.F. and Blattner, F.R. (1980). *Science* **209**, 1353–1360.
Tucker, P.W., Slighton, J.L. and Blattner, F.R. (1981). *Proc. Natl Acad. Sci. USA* **78**, 7684–7688.
Tutter, A. and Riblet, R. (1988). *Curr. Top. Microbiol. Immunol.* **137**, 107–115.
Tutter, A., Brodeur, P., Shlomchik, M. and Riblet, R. (1991). *J. Immunol.* **147**, 3215–3223.
Tyler, B.M., Cowman, A.F., Genondakis, S.D., Adams, J.M. and Bernerd, O. (1982). *Proc. Natl Acad. Sci. USA* **79**, 2008–2012.
Ueda, S., Nakai, S., Nishida, Y., Hisajima, H. and Honjo, T. (1982). *EMBO J.* **1**, 1539–1544.
van Dijk, K.W., Mortari, F., Kirkham, P.M., Schroeder, H.W. and Milner, E.C.B. (1993). *Eur. J. Immunol.* **23**, 832–839.
Walter, G., Tomlinson, I.M., Cook, G.P., Winter, G., Rabbitts, T.H. and Dear, P.H. (1993). *Nucleic Acids Res.* **21**, 4524–4529.

Walter, M.A., Surti, U., Hofker, M.H. and Cox, D.W. (1990). *EMBO J.* **9**, 3303–3313.
Walter, M.A., Dosch, H.M. and Cox, D.W. (1991a). *J. Exp. Med.* **174**, 335–349.
Walter, M.A., Gibson, W.T., Ebers, G.C. and Cox, D.W. (1991b). *J. Clin. Invest.* **87**, 1266–1273.
Wels, J.A., Word, C.J., Rimm, D., Der-Balan, G.P., Martinez, H.M., Tucker, P.W. and Blattner, F.R. (1984). *EMBO J.* **3**, 2041–2046.
White, M.B., Shen, A.W., Word, J.C., Tucker, P.W. and Blattner, F.R. (1985). *Science* **228**, 733–737.
Winter, E., Radbruch, A. and Krawinkel, U. (1985). *EMBO J.*, **4**, 2861–2867.
Wood, C. and Tonegawa, S. (1983). *Proc. Natl Acad. Sci. USA* **80**, 3030–3034.
Word, J.C., Munshinski, J.F. and Tucker, P.W. (1983). *EMBO J.* **2**, 887–898.
Wu, G.E. and Paige, C.J. (1986). *EMBO J.* **5**, 3475–3481.
Wu, G.E. and Paige, C.J. (1988). *J. Exp. Med.* **167**, 1499–1504.
Yamawaki-Kataoka, Y., Kataoka, T., Takahashi, N., Obata, M. and Honjo, T. (1980). *Nature* **283**, 786–789.
Yamawaki-Kataoka, Y., Miyata, T. and Honjo, T. (1981). *Nucleic Acids Res.* **9**, 1365–1381.
Yamawaki-Kataoka, Y., Nakai, S., Miyata, T. and Honjo, T. (1982). *Proc. Natl Acad. Sci. USA* **79**, 2623–2627.
Yancopoulos, G.D., Desiderio, S.V., Pasking, M., Kearney, J.F., Baltimore, D. and Alt, F.W. (1984). *Nature* **311**, 727–733.
Yancopoulos, G.D., Maylnn, B.A. and Alt, F.W. (1988). *J. Exp. Med.* **168**, 417–435.
Yang, P.-H., Olsen, N.J., Siminovitch, K.A., Olee, T., Kozin, F., Carson, D.A. and Chen, P.P. (1990). *Proc. Natl Acad. Sci. USA* **88**, 7907–7911.
Zong, S.Q., Nakai, S., Matsuda, F., Lee, K.H. and Honjo, T. (1988). *Immunol. Lett.* **17**, 329–334.
Zouali, M. (1992). *Immunol. Rev.* **128**, 73–99.

Genetic and Immunological Properties of Phage-Displayed Human Anti-Rh(D) Antibodies: Implications for Rh(D) Epitope Topology

By Tylis Y. Chang and Don L. Siegel

Understanding anti-Rh(D) antibodies on a molecular level would facilitate the genetic analysis of the human immune response to Rh(D), lead to the design of therapeutically useful reagents that modulate antibody binding, and provide relevant information regarding the structural organization of Rh(D) epitopes. Previously, we described a Fab/phage display-based method for producing a large array of anti-Rh(D) antibodies from the peripheral blood lymphocytes of a single alloimmunized donor. In the current study, we present a detailed analysis of 83 randomly selected clones. Sequence analysis showed the presence of 28 unique γ_1 heavy chain and 41 unique light chain gene segments. These paired to produce 53 unique Fabs that had specificity for at least half of the major Rh(D) epitopes. Surprisingly, despite this diversity, only 4 closely related heavy chain germline genes were used (VH3-30, VH3-30.3, VH3-33, and VH3-21). Similarly, nearly all V_κ light chains (15/18) were derived from one germline gene (DPK9). λ light chains showed a more diverse V_L gene usage, but all (23/23) used the identical $J_\lambda 2$ gene. Several Fabs that differed in epitope specificity used identical heavy chains but different light chains. In particular, 2 such clones differed by only 3 residues, which resulted in a change from epD2 to epD3 specificity. These results suggest a model in which footprints of anti-Rh(D) antibodies are essentially identical to one another, and Rh(D) epitopes, as classically defined by panels of Rh(D) variant cells, are not discrete entities. Furthermore, these data imply that the epitope specificity of an anti-Rh(D) antibody can change during the course of somatic mutation. From a clinical perspective, this process, which we term epitope migration, has significance for the design of agents that modulate antibody production and for the creation of mimetics that block antibody binding in the settings of transfusion reactions and hemolytic disease of the newborn.

CLINICALLY, THE HUMAN Rh(D) antigen is the most important red blood cell (RBC) membrane protein in transfusion medicine. The alloimmune response against Rh(D) produces high-affinity IgG antibodies that cause hemolytic transfusion reactions and hemolytic disease of the newborn (HDN). The prophylactic use of Rh(D)-immune globulin in pregnant Rh(D)-negative women has been a major advance in the prevention of HDN,[1] yet the mechanism by which the drug exerts its immune modulatory effect is not well understood.[2] Monoclonal antibodies (MoAbs) derived from the B cells of Rh(D)-immune globulin donors have defined several dozen Rh(D) epitopes[3]; paradoxically, the Rh(D) antigen, an approximately 30-kD transmembrane protein, has minimal extracellular mass and presents a very limited surface area for epitope expression.[4-9] The molecular cloning of large repertoires of anti-Rh(D) antibodies would help reconcile these observations. In addition, it would facilitate the rational development of recombinant formulations of Rh(D)-immune globulin and aid in the design of therapeutic agents that block antibody binding. Furthermore, the comprehensive genetic analysis of anti-Rh(D) antibodies within a given alloimmunized individual would serve as a paradigm for human immune repertoire development, an area in which limited information is currently available.

Previously, no more than 8 IgG anti-Rh(D) human MoAbs have been derived from a single individual.[10] The primary challenge in studying the Rh(D) immune response has been technical difficulties inherent in human B-cell immortalization. Epstein-Barr virus (EBV) transformation results in relatively low transformation efficiencies[11] that can undergo a decline in antibody production,[12-15] whereas cell fusion methods have been hampered by the lack of good fusion partners.[16,17] More recently, molecular approaches have been developed that bypass the need for cell transformation.[18-20] Conceptually, these techniques, referred to as repertoire cloning or Fab/phage display, seek to immortalize Ig mRNA rather than the B cells from which they were derived. In an earlier report, our laboratory adapted these techniques for isolating Fab/phage antibodies directed against conformation-dependent antigens expressed on cell surfaces.[21] Using intact human RBCs, we isolated highly diverse $\gamma_1\kappa$ and $\gamma_1\lambda$ Fab/phage libraries against the Rh(D) antigen from the B cells of a single Rh(D)-immune globulin donor.[22]

In the following report, we present a detailed genetic and serological analysis of 53 unique anti-Rh(D) antibodies derived from 83 randomly chosen clones. The results complement previous reports on the genetic and biochemical makeup of monoclonal anti-Rh(D) antibodies derived from multiple donors.[10,23-25] Significantly, our data also demonstrate extensive genetic homology between antibodies directed against different Rh(D) epitopes. We provide evidence that antibodies directed against different epitopes can be clonally related. Finally, we suggest a model that reconciles the serological diversity of anti-Rh(D) antibodies with the topological constraints imposed by the Rh(D) antigen.

From the Blood Bank/Transfusion Medicine Section, Department of Pathology & Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA.

Submitted September 29, 1997; accepted November 20, 1997.

Supported in part by March of Dimes Birth Defects Foundation Basil O'Connor (Grant No. 5-FY94-0787) and Clinical Research (Grant No. 6-FY96-0367) Awards (D.L.S.), by a National Institutes of Health Specialized Center of Research (SCOR) in Transfusion Medicine and Biology Award (P50-HL54516; to D.L.S.), and by a grant from the National Blood Foundation (T.Y.C.).

Address reprint requests to Don L. Siegel, PhD, MD, Department of Pathology & Laboratory Medicine, 6-55 Founders Pavilion, Hospital of the University of Pennsylvania, 3400 Spruce St, Philadelphia, PA 19104.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. section 1734 solely to indicate this fact.

0006-4971/98/9108-0020$3.00/0

MATERIALS AND METHODS

Production of Monoclonal Anti-Rh(D) Phage-Displayed and Soluble Fab Molecules

Methods for the isolation of human anti-Rh(D)–specific antibodies from $\gamma_1\kappa$ and $\gamma_1\lambda$ Fab/phage display libraries using the pComb3H phagemid vector and a cell-surface panning protocol have been previously published.[22] Soluble anti-Rh(D) Fab preparations for inhibition studies were produced from bacterial cultures transfected with plasmid DNA from which the M13 gene III coat protein sequence had been excised as described.[21,26] Cultures were grown by shaking at 300 RPM at 37°C in superbroth (30 g/L tryptone, 20 g/L yeast, 10 g/L MOPS, pH 7.00) containing 20 mmol/L $MgCl_2$ and 50 μg/mL carbenicillin to an OD_{600} of 0.5. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 1 mmol/L and cultures were shaken overnight at 30°C. Bacterial pellets were harvested and resuspended in 1/50th of the initial culture volume with osmotic shock buffer (500 mmol/L sucrose, 1 mmol/L EDTA, 100 mmol/L Tris, pH 8.00), incubated for 30 minutes at 4°C, and centrifuged at 16,000*g* for 15 minutes at 4°C. Fab-containing supernatants were dialyzed against phosphate-buffered saline (PBS) and used in agglutination experiments without further purification.

Anti-Rh(D) Antibody Binding Assays

The binding of anti-Rh(D) Fab/phage or soluble Fab molecules to normal or partial Rh(D) antigens was assessed by indirect agglutination assays as described.[21,22] Briefly, 100 μL aliquots of phage-displayed Fabs or soluble Fabs were incubated with 50 μL of a 3% suspension of RBCs. After 1 hour of incubation at 37°C, the RBCs were washed three times with 2 mL of cold PBS to remove unbound antibody. The resulting RBC pellets were resuspended in 100 μL of a 10 μg/mL solution of sheep anti-M13 antibody (5 Prime → 3 Prime, Boulder, CO) for Fab/phage experiments or goat antihuman κ or λ light chain antibody (Tago, Burlingame, CA) for $\gamma_1\kappa$ or $\gamma_1\lambda$ soluble Fab experiments, respectively. The RBC suspensions were transferred to the round-bottomed wells of a 96-well microplate and left undisturbed for 2 hours. Negative reactions show sharp approximately 2-mm diameter RBC spots, whereas the RBCs in agglutinated wells form a thin carpet coating the entire floor of the well.[22] Agglutination titers for recombinant antibodies were determined by performing serial twofold dilutions in 1% bovine serum albumin (BSA)/PBS. Typically, Fab/phage had agglutination titers of 1/1,024 to 1/2,048 (where neat is defined as 5×10^{12} tfu/mL),[22] and soluble Fabs had agglutination titers of 1/64 to 1/128 when prepared as described above.

For determining Rh(D) epitope specificity for anti-Rh(D) Fab/phage antibodies, the following reference Rh(D) variant cells were obtained from the MRC Blood Group Unit (London, UK), The New York Blood Center (New York, NY), or Gamma Biologicals, Inc (Houston, TX): $O/D^{III a}Cce$, G positive; $B/D^{III c}Cce$; $A/D^{IVa}ce$; $A/D^{IVa}ce$; $O/D^{IVa}ce$; $O/D^{IVb}Cce$; $B/D^{IVb}Cce$, Go^a negative, Rh32 negative; $O/D^{Va}Cce$; $O/D^{Va}cEe$, D^w positive; $O/D^{VI}Cce$; $B/D^{VI}Cce$; $AB/D^{VI}Cce$; $A/D^{VI}cEe$; $O/D^{VII}Ccc$; and $O/D^{VII}Cce$. Each Fab/phage antibody was tested on at least three separate occasions against at least two different examples of each variant cell type, and identical epitope assignments were obtained each time. For antibodies that demonstrated previously undescribed patterns of reactivity or repeatedly weak reactivity against one type of cell (see the Results), monoclonal Fab/phage were prepared on a least four separate occasions to verify the patterns of reactivity.

For inhibition studies, the ability of antibodies with different Rh(D) epitope specificities to compete with each other for binding was assessed by preparing stocks of each clone in both a soluble Fab form and a phage-displayed form. Pairwise combinations of soluble Fabs and Fab/phage were prepared and added to Rh(D)-positive RBCs. The resulting incubation mixes comprised 50 μL of a 3% suspension of RBCs, 100 μL of undiluted soluble Fab, and 100 μL of Fab/phage diluted to its highest agglutinating titer. After 1 hour of incubation at 37°C, RBCs were washed, resuspended in anti-M13 antibody, and placed in microplate wells as described above. That the amount of soluble Fab present in an incubation mixture was sufficient to compete away a Fab/phage that shared the same binding site was determined by verifying that each soluble Fab preparation could block its own Fab/phage (see the Results).

Inhibition experiments were also performed using pairwise combinations of soluble Fabs instead of soluble Fab and Fab/phage combinations. In this type of experiment, pairs of soluble Fabs specific for different epitopes were chosen such that one Fab contained a λ light chain and the other a κ light chain. Incubations with RBCs were performed with one Fab in excess and the other in limiting amounts. Blocking of the latter antibody was assessed using a secondary antibody (anti-λ or anti-κ) specific for its light chain isotype.

Nucleotide Sequencing and Analysis

Plasmid DNA for sequencing was prepared using the Qiawell system (Qiagen, Chatsworth, CA). Double-stranded DNA was sequenced using light chain or heavy chain Ig constant region reverse primers or a set of unique pComb3H vector primers that anneal 5′ to the respective Ig chain[26,27] and automated fluorescence sequencing (Applied Biosystems, Foster City, CA; DNA Sequencing Facility, University of Pennsylvania Department of Genetics and Cancer Center, Philadelphia, PA). Sequence analysis and variable region germline assignments were performed using DNAplot[28] and the V Base Directory of Human V Gene Sequences (March 1997 update).[29] Germline assignments were corroborated with the MacVector (v. 6.0) software package (Oxford Molecular Group, Oxford, UK) against the same database. Multiple sequence alignments and predictions of isoelectric point were calculated using the Pileup and Isoelectric programs of the GCG software package (v. 8.0.1; GCG, Madison, WI). Statistical analysis was performed with Statview (Abacus Concepts, Berkeley, CA).

Because of the large number of heavy and light chain sequences (N = 69), only alignments of the predicted amino acid sequences are presented. Nucleotide sequences of all clones are available in Genbank.

RESULTS

Sequence Analysis of Anti-Rh(D) Heavy and Light Chains

We previously reported on the use of Fab/phage display and cell-surface panning to isolate a large array of anti-Rh(D) antibodies from the peripheral blood lymphocytes of a single hyperimmunized donor.[22,30] Separate $\gamma_1\kappa$ and $\gamma_1\lambda$ Fab/phage display libraries had been constructed and contained 7×10^7 and 3×10^8 independent transformants, respectively, based on electroporation efficiency. Each library was panned independently using a simultaneous positive/negative selection strategy with magnetically labeled Rh(D)-positive RBCs and unmodified Rh(D)-negative RBCs as described. After two rounds of panning, 32 of 36 $\gamma_1\lambda$ and 15 of 15 $\gamma_1\kappa$ randomly chosen clones were positive for anti-Rh(D) activity. After the third round of panning, 24 of 24 $\gamma_1\lambda$ and 12 of 12 $\gamma_1\kappa$ clones were positive. Nucleotide sequencing of the 83 positive clones showed a total of 28 unique heavy and 41 unique light chains. Because of combinatorial effects during phage display library construction, heavy and light chain gene segments paired to produce 53 unique Fab antibodies.[22]

Anti-Rh(D) heavy chains. All of the heavy chain sequences used V_HIII family-encoded gene products (Figs 1 and 2).

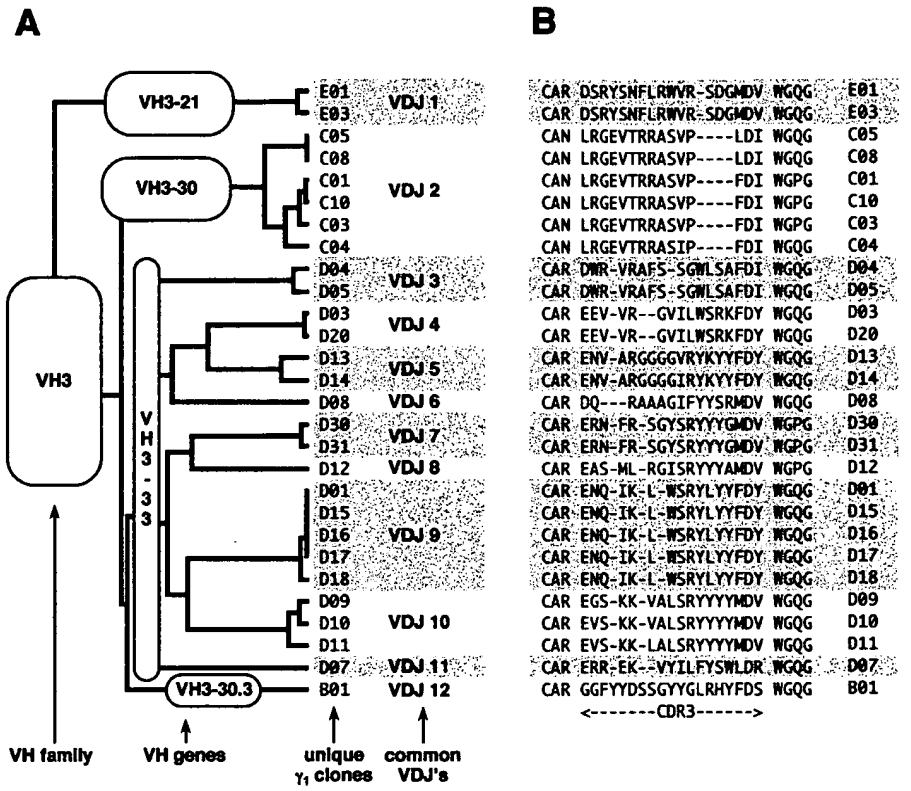


Fig 1. (A) Dendrogram and (B) CDR3 alignment of anti-Rh(D) heavy chains. The 28 unique heavy chain clones are organized by V_H family, V_H germline gene, and VDJ rearrangement. Each heavy chain clone is identified by a numeral preceded by a letter (B through E) that denotes its germline gene. The 28 heavy chains comprised 12 distinct VDJ regions, designated VDJ1 through VDJ12. Clones with identical VDJ joins putatively result from intraclonal diversity of 12 original B lymphocytes.

Several sequences shared identical VDJ joining regions, and 12 unique VDJ rearrangements were identified and designated VDJ1 through VDJ12. Alignment of these sequences against the V Base Directory of Human V Gene Sequences[29] showed that only four V_HIII genes were used by these antibodies: VH3-21, VH 3-30, VH 3-33, and VH 3-30.3. VH3-21 was used by 1 of the 12 VDJs and 2 of the 28 clones; VH3-30 by 1 VDJ and 6 clones; VH3-33 by 9 VDJs and 19 clones; and VH3-30.3 by 1 VDJ and 1 clone. Interestingly, VH3-30, VH3-33, and VH3-30.3 comprise a set of closely related genes (>98% homology; Fig 2B) and their next nearest neighbor, VH3-07, is only 90% homologous (Fig 2C). Hereafter, these three genes are referred to as the VH3-33 superspecies. Heavy chain E1 differed from VH3-21 by 6 mutations and from VH3-48 by 10 mutations; hence, it was assigned to the former germline gene. Because there were no common mutations among the VH3-33 clones, it is highly probable that the donor possessed the VH3-33 germline gene. However, we could not formally rule out gene duplication with allelic variants of VH3-33 or the existence of variant alleles of the other germline genes in the donor. The isolation of clones sharing multiple VDJ joining regions argues that cloning artifacts cannot account for the V_H gene restrictions observed.

Neither J_H nor D segments showed restriction. At least 9 different D segments were used and J_H gene use comprised J_H6 (5 VDJs and 9 clones), J_H4 (4 VDJs and 10 clones), J_H3 (2 VDJs and 8 clones), and J_H5 (1 VDJ and 1 clone). All four V_H genes were Chothia class 1-3,[31] and the CDR3s showed a narrow range of length from 15 to 19 residues.

Because rearranged heavy chain genes demonstrate extensive diversity, clones sharing identical VDJ rearrangements are generally considered to have arisen from the same clone. Based on nucleotide alignment with the germline genes (data not shown), an ontogeny tree was constructed for the 12 VDJs and 28 clones (Fig 3). By using the most parsimonious mutation scheme (ie, postulating the minimum number of mutations), putative intermediate antibodies were derived for several of the VDJs and were designated Ca, Cb, Da, Db, and Dc (Figs 2A and 3). Compared with the isolated heavy chain clones, which displayed between 6 and 23 nucleotide differences from their germline counterparts, these putative intermediates had between 3 and 12 mutations from germline. Based on the ontogeny tree, the number of independent mutations could be tabulated among the clones. The most commonly mutated residues were 52a and 58 (7 independent mutations), followed by residues 30, 31, and 50 (6 mutations) and residue 55 (5 mutations). In the VH3-33 superspecies, residues 52a and 58 in CDR2 are tyrosines and residue 52a was mutated to phenylalanine in 6 of the 11 VDJs derived from VH3-33 superspecies V_H genes. Mutations at residue 58 comprised glutamate (3), aspartate (2), histidine (1), and asparagine (1). The AGY serines at residues 30, 31, and 55 were mutated to a number of different amino acids, although the AGY serine at 82b was conserved in all clones. The valine at residue 50 in the VH3-33 superspecies also had a diverse set of mutations. This distribution of hot spots is similar to that seen with nonproductive rearrangements as previously reported by Dörner et al.[32]

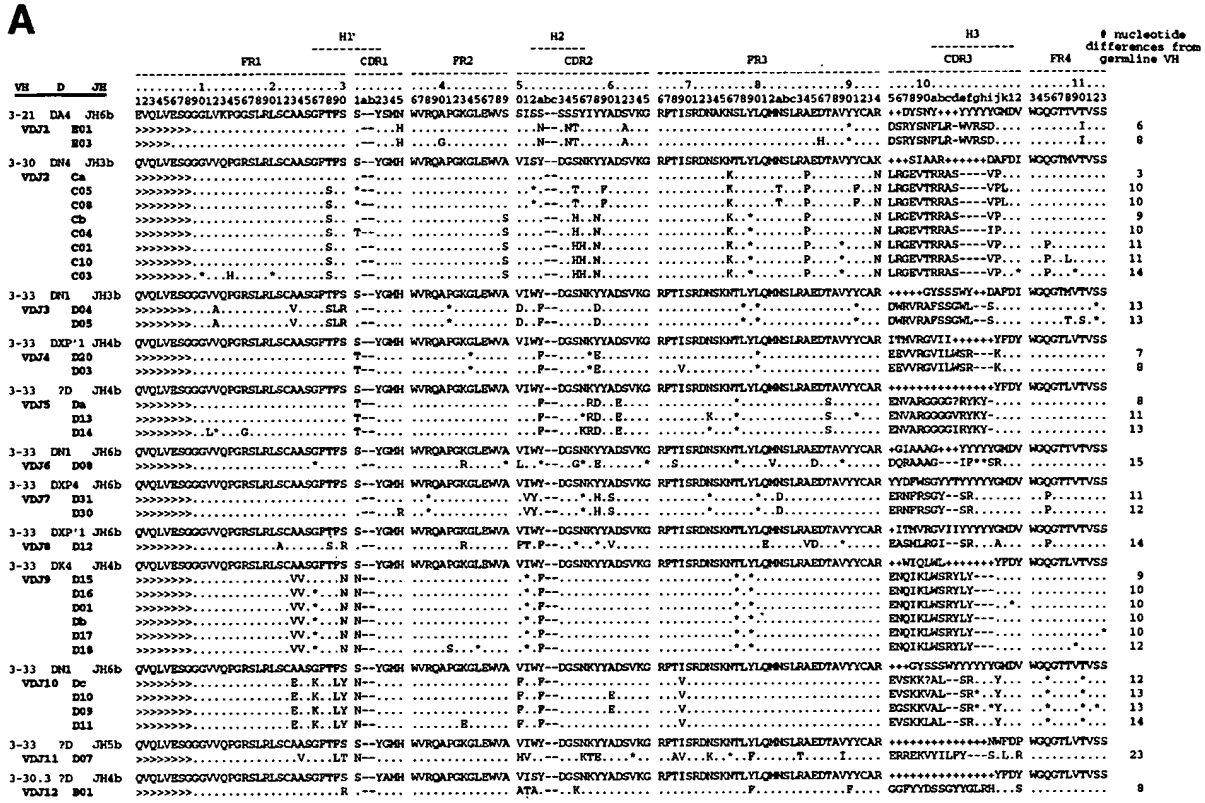


Fig 2. (A) Alignment of anti-Rh(D) heavy chains to their nearest germline V, D, and J genes. Also shown are the putative intermediate heavy chain sequences (Ca, Cb, Da, Db, and Dc; see text and Fig 3). The number of nucleotide differences from a germline V_H is tabulated to the right of each sequence. In general, D segments showed poor homology with known D genes, so mutations were not scored in these regions. Key: Replacement mutations indicated with letters, silent mutations as "*", identities as ".", and insertions as "-". Sequences derived from the 5′ V_H primers used in library construction[22] are marked as ">". CDR region designations are determined as per Kabat[59]; numbering and H region designations per Chothia et al.[31] (B) Alignment of the four VH3 genes used by anti-Rh(D) heavy chains and (C) dendrogram of all human VH3 family germline genes shows relatedness of VH3-21, VH3-30, VH3-33, and VH3-30.3 and the surprising restriction in V_H gene usage. Note that the VH3-30.5 gene is present in only certain haplotypes and is identical to VH3-30.[60] Genbank accession numbers for anti-Rh(D) heavy chains are listed in the Appendix.

Anti-Rh(D) light chains. Seventeen of the 18 κ light chains were from the $V_κ$I family and the remaining light chain originated from a $V_κ$II family member germline gene (Fig 4). Only 4 $V_κ$ germline genes were used (15 clones were derived from DPK9 alone), and the κ light chain clones had between 1 and 49 mutations from their corresponding $V_κ$ germline genes. All 5 of the known $J_κ$ genes were used and were each joined to the DPK9 gene in one or more clones. Because the light chains showed considerably less diversity in their joining regions than the heavy chains, it was difficult to assign common clonal

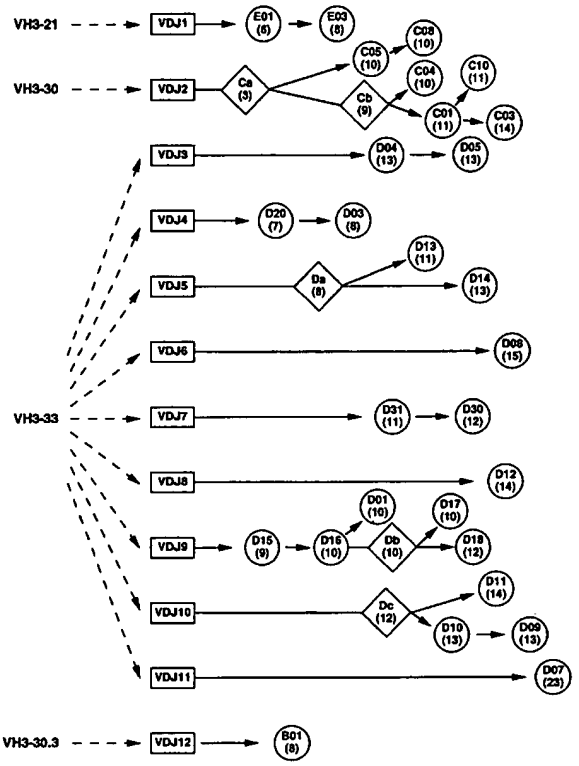


Fig 3. Ontogenic tree of anti-Rh(D) heavy chains constructed using nucleotide alignment data. Circles represent isolated and sequenced clones and diamonds represent putative intermediates (see Fig 2A). The number of nucleotide mutations from its germline V_H gene is shown in parentheses below the clone name. The distance along the horizontal axis represents the degree of mutation (including J segments) within the constraints of the diagram.

origins. However, an ontogeny tree was constructed by grouping common V and J gene segments along with common mutations (data not shown). Based on this analysis, the 18 κ chains comprised at least 10 different recombination events.

λ light chains were restricted by their J_λ gene usage but showed no restriction in their use of V_λ genes (Fig 5). The 23 λ light chains all used the J_λ2Vasicek gene but were derived from V_λI (12 clones), V_λIII (5), V_λVII (3), V_λII (2), and V_λIV (1) family genes. The number of mutations ranged from 2 to 41 from the nearest germline V_λ gene. Based on common joining regions and mutations, these 23 λ light chains were derived from at least 13 different B cells.

Assessment of the Diversity of the Unpanned Libraries

To determine whether the apparent restriction in gene usage of the anti-Rh(D) antibodies could have been due to preselection factors (ie, cloning artifacts), we assessed the diversity of the unpanned $\gamma_1\kappa$ and $\gamma_1\lambda$ Fab/phage libraries. By sequencing 39 randomly picked clones, we determined that there were no duplicate heavy or light chain sequences and that there was significant heterogeneity in V gene family representation before selection (Fig 6). In fact, the variable region gene family distribution was not unlike that found by other investigators for IgG-secreting lymphocytes in adult peripheral blood.[33] Furthermore, of the 14 V_HIII-encoded negative clones, only one used a VH3-33 superspecies germline gene (VH3-30.3); the other 13 were encoded by VH3-07 (3), 3-09 (2), 3-15 (2), 3-48 (2), 3-72 (2), 3-23 (1), and DP-58 (1). Therefore, the restriction of the 83 anti-Rh(D) clones to the VH3-33, 3-30, 3-30.3, and 3-21 genes is significant and not a result of skewed representation of certain germline genes within the originally constructed $\gamma_1\kappa$ and $\gamma_1\lambda$ Fab/phage libraries.

Heavy and Light Chain Contribution to Rh(D) Epitope Specificity

Because of the conformational dependency of Rh(D) antigenicity, Rh(D) epitopes have been classically defined through the use of RBCs obtained from rare individuals whose cells appear to produce Rh(D) antigens lacking certain epitopes.[34] Examining the pattern of agglutination of a particular anti-Rh(D) MoAb with such sets of partial Rh(D) RBCs enables one to categorize that antibody's fine specificity.

Monoclonal Fab/phage preparations were prepared in triplicate for each of the 53 anti-Rh(D) clones and tested against a panel of Rh(D) category cells IIIa/c, IVa, IVb, Va, VI, and VII. This panel of cells can differentiate between the Rh(D) epitope specificities as described by Lomas et al[6] (designated epitopes epD1, epD2, epD3, epD4, epD5, and epD6/7). Agglutination experiments with the Fab/phage clones showed five different patterns of reactivity, including a new pattern that had not been described in the original study by Lomas et al[6] or in the more recently described 9, 30, or 37 epitope systems (Figs 7 and 8).[3,35] Although nearly all Fab/phage gave unequivocal agglutination reactions, a few antibodies gave repeatedly weak patterns of reactivity against one of the panel cells. For these reactions, monoclonal Fab/phage were prepared on at least four separate occasions to verify the patterns of reactivity.

The most commonly recognized epitope was epD6/7, against which 13 clones were directed. Interestingly, monoclonal anti-Rh(D) clones isolated using conventional tissue culture methods are most often specific for epD6/7.[34] epD2, epD1, and epD3 were recognized by 10, 7, and 2 clones, respectively. Six clones agglutinated cells of categories IIIa/c, IVa, and VII, but not of categories IVb, Va, and VI, and were designated anti-epDX. This pattern is identical to epD1, except that the IVa cell is agglutinated. Three clones gave intermediate reactions with cell IVa, but otherwise showed patterns consistent with epDX or epD1. These clones were designated epDX^1 or epD1^X, depending on whether this reactivity against cell IVa was stronger or weaker, respectively (Fig 8). Similarly, reaction patterns for epD1 and epD2 differ by a positive reaction with the category Va cell; therefore, one clone was given epD2^1 specificity because it gave only moderate reactivity against that cell. Such variable reactions against one or more partial Rh(D) cells have been observed for anti-Rh(D) MoAbs produced through conventional tissue culture methods.[36]

Because of the reassortment of heavy and light chain gene segments that occurs during the construction of a phage display library, a number of clones were isolated that shared either a heavy (eg, E1) or light (eg, M3) chain sequence (Fig 8). Some heavy chains were found to have paired with both κ and λ light chains (eg, C1, D20), and each demonstrated anti-Rh(D)

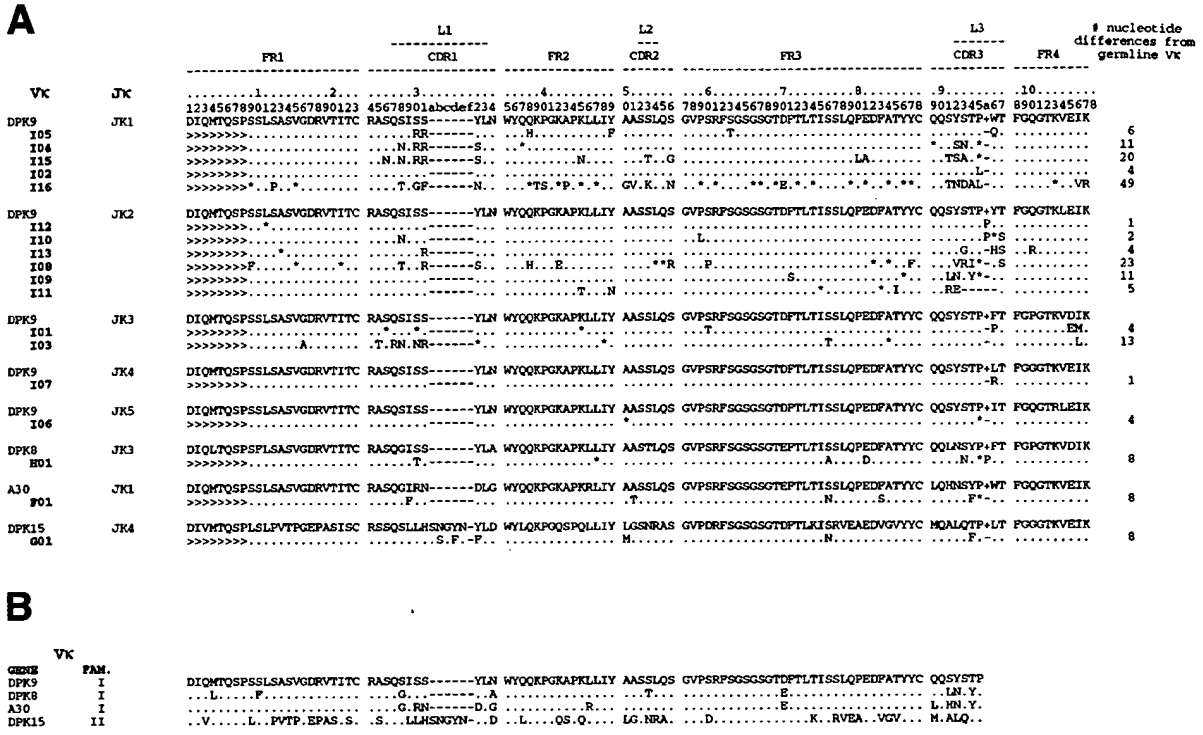


Fig 4. (A) Alignment of anti-Rh(D) κ light chains to their nearest germline V and J genes shows predominance of DPK-9 usage from the $V_κI$ family. Nomenclature for clones is similar to that for heavy chains but uses the letters F through I. (B) Alignment of the four $V_κ$ genes used by anti-Rh(D) light chains. The key is the same as that used in Fig 2A. Genbank accession numbers for anti-Rh(D) κ light chains are listed in the Appendix.

specificity. Interestingly, some heavy chains (eg, E1, D12) mapped to different epitopes depending on the light chains with which they were paired. In particular, the light chains of two such clones, E1/M2 and E1/M3, differed by only 3 amino acid residues (Fig 5) and these differences appear to confer specificity for epD2 versus epD3.

Inhibition Studies

To investigate the topological relationships among the Rh(D) epitopes, inhibition studies were performed. Previous work by Gorick et al[37] using pairs of unlabeled and ^{125}I-labeled anti-Rh(D) MoAbs demonstrated that antibodies to at least 3 different Rh(D) epitopes (subsequently identified as epD1, D6, and D7)[6] could inhibit one another. We have confirmed and extended these findings using recombinant antibodies to 5 Rh(D) epitopes (Fig 9). In one series of experiments, we exploited the ability to express each antibody in both a soluble Fab as well as phage-displayed form to ask whether a soluble Fab against one epitope would inhibit the agglutination induced by an Fab/phage directed against a different epitope. Reciprocal pairs of soluble Fab and Fab/phage specific for epD1, epD2, epD3, epD6/7, and epDX were tested. All 10 combinations showed mutual inhibition patterns (shown in Fig 9A for an anti-epD3/anti-epD6/7 combination). To show that this inhibition was not due to nonspecific factors, a control with an irrelevant RBC-binding recombinant antibody (an anti-blood group B antibody) was performed (Fig 9B). That sufficient inhibitory amounts of soluble Fab was present were first verified by demonstrating that each soluble Fab could inhibit its own Fab/phage (Fig 9A and B; samples on diagonal). Similar results were obtained using pairs of soluble Fabs which differed in their light chain isotype composition (Fig 9C).

Isoelectric Point (pI) Analysis of Anti-Rh(D) Antibodies

The restriction in V_H germline gene usage to only four V_HIII family members was intriguing in light of their ability to confer specificity to a number of Rh(D) epitopes. As suggested by Boucher et al,[10] V_H germline gene segments used to encode anti-Rh(D) antibodies are among the most cationic segments available in the human V_H repertoire that may be used to account for the relatively high pI of polyclonal anti-Rh(D)–containing antisera.[38,39] Although the cationic nature of the antibodies may be important for binding to Rh(D), it has also been suggested that a constitutive net positive charge may be necessary to permeate the highly negative RBC ζ potential, thus permitting antibody to contact antigen.[34] In either case, analysis of the predicted pI for the 28 heavy chains and 41 light chains isolated here showed an interesting phenomenon for the heavy versus light chains. Using the pI interval scale of Boucher et al,[10] the average pI of the 4 germline V_H segments used to encode the 28 heavy chains is high (9.87 ± 0.15) and significantly higher than that of 39 randomly picked, non-Rh(D) binding clones from the original unpanned libraries (9.24 ± 0.80, $P < 10^{-5}$). Similar to the results of Boucher et al,[10] the addition of D and J_H segments and the introduction of somatic mutation did not significantly change the pI of the average

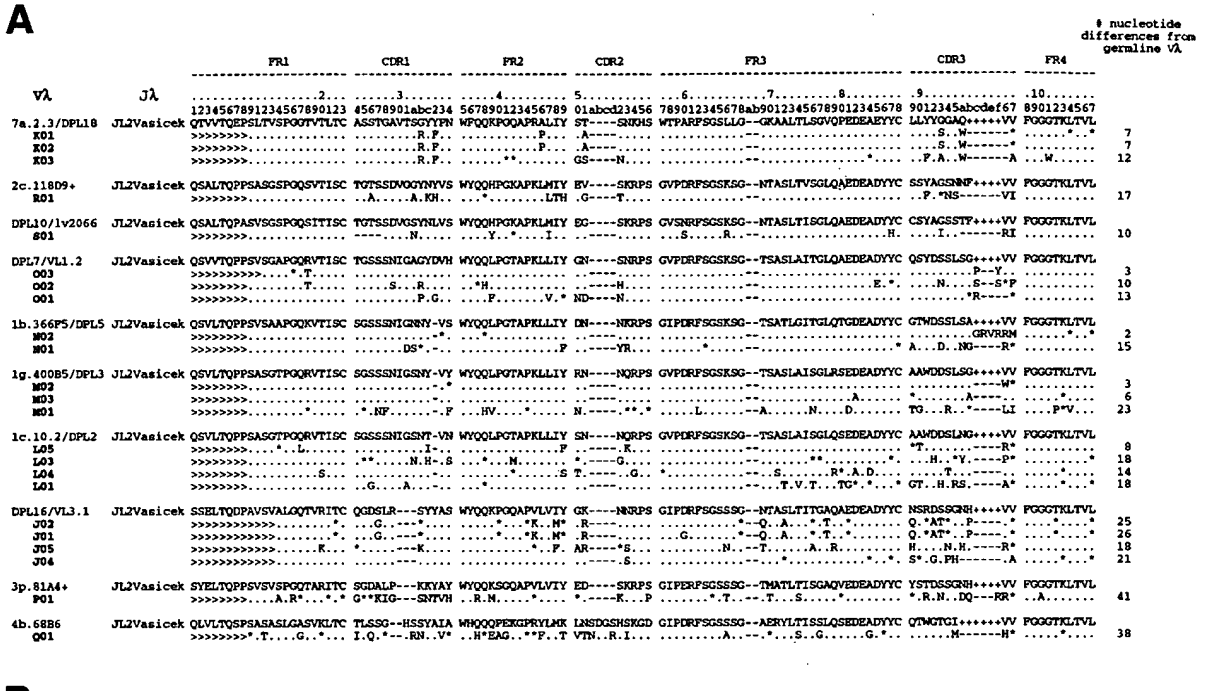


Fig 5. (A) Alignment of anti-Rh(D) λ light chains to their nearest germline V and J genes and (B) alignment of the 10 V_λ germline genes used shows the use of a diverse set of variable region genes derived from multiple families. However, all of the clones use the identical J_λ gene segment. Nomenclature for the clones is similar to that for heavy chains but uses the letters J through S. The key is the same as that used in Fig 2A. Genbank accession numbers for anti-Rh(D) λ light chains are listed in the Appendix.

anti-Rh(D) heavy chain (9.81 ± 0.33, $P < .37$). However, for the light chains, the average pI of their germline counterparts was not cationic, but the light chains became so through the addition of J_L segments and somatic mutation. Overall, for all 18 κ and 23 λ light chains, paired *t*-test analyses before and after somatic mutation showed a significant increase in net positive charge when comparing germline V_L (6.63 ± 1.47) with expressed V_L (7.28 ± 1.51, $P < 10^{-3}$) or germline V_LJ_L (7.43 ±

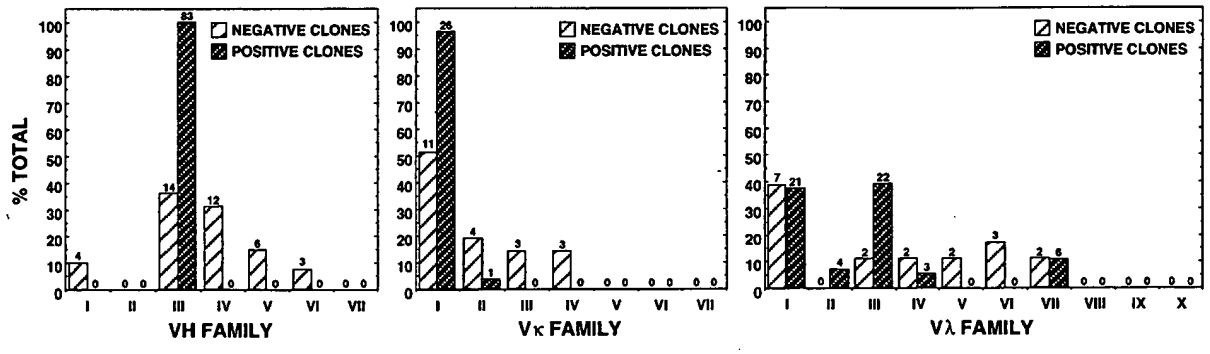


Fig 6. Comparison of variable region gene family usage for anti-Rh(D)–specific clones with those used by randomly picked, non-Rh(D)-binding clones from original $\gamma_1\kappa$ and $\gamma_1\lambda$ unselected libraries. Lightly hatched bars reveal heterogeneity in V_H (left panel), V_κ (middle panel), and V_λ (right panel) family representation before selection for anti-Rh(D) specificity. Numbers above bars represent absolute number of clones in that group.

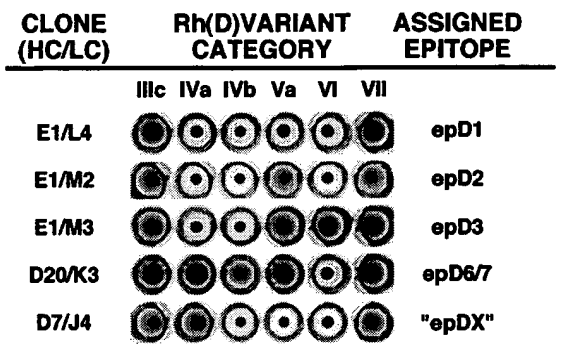


Fig 7. Determination of the Rh(D) binding epitope of anti-Rh(D) Fab/phage clones. The fine specificities of monoclonal Fab/phage clones were determined by their ability to agglutinate members of a panel of six Rh(D) variant RBCs. Shown are the five different agglutination patterns obtained from screening all of the 53 Fab/phage clones. The particular clones shown are identified by their unique heavy chain/light chain pairings using the nomenclature defined in Figs 1, 4, and 5. For E1/M3, reactivity with additional Rh(D) variant cells would be required to distinguish its specificity for epD3 versus epD9.[3] Rh(D) epitope assignments are as per Lomas et al.[6] Note that inclusion of the category IVb cell (not available in our previous study)[22] permits the identification of a new epitope designated epDX (see text).

1.47) with expressed V_LJ_L (8.55 ± 1.35, $P < 10^{-7}$). There was no significant increase in a similar analysis of 16 non-Rh(D) binding clones ($P < .59$ and $P < .19$, respectively). Examination of the light chain sequences (Figs 4 and 5) showed that this increase in pI resulted from mutations that not only introduced positively charged residues, but also eliminated some negatively charged residues. There were 31 such events, 29 (91%) of which occurred in the light chain CDR regions.

DISCUSSION

Conventional and Phage-Displayed Anti-Rh(D) MoAbs

Because of differences in methodology, we were interested in comparing our phage-display–derived anti-Rh(D) clones with those produced by conventional tissue culture techniques (EBV transformation and cell fusion). Despite the relatively small number of previously published sequences for IgG anti-Rh(D) antibodies (N = 21) and the fact that they were derived from over 10 different donors,[10,23-25] there was surprisingly good correlation between the two groups (Table 1). Both cohorts demonstrated a predominance of V_HIII-family encoded germline genes, particularly those of the VH3-33 superspecies. CDR3 regions showed similar lengths, ranging from 15 to 19 residues for Fab/phage antibodies and 16 to 20 for conventional monoclonals, although one heterohybridoma was an outlier with a CDR3 length of 28 residues. κ light chains were biased towards $V_κ1$ family members and λ light chains demonstrated the preferential use of the $J_λ2$Vasicek gene. The only qualitative discrepancy was in $V_λ$ family usage, where Fab/phage clones demonstrated a slight preference for $V_λI$ versus $V_λIII$ family members for conventional monoclonals. However, in both cohorts, DPL16 was used more often than any other λ light chain gene.

It has been suggested in the literature that the VH4-34 (VH4.21) germline gene, a gene used by many autoantibodies and cold agglutinins,[40-42] may play an important role in the immune response to Rh(D).[43] However, these conclusions arose from the analysis of *IgM* monoclonals and only 2 of the 21 published anti-Rh(D) IgG sequences used VH4-34.[25] In a related series of experiments, we pooled aliquots of the $γ_1κ$ and $γ_1λ$ libraries obtained after the second and third rounds of selection and then panned them against the VH4-34 specific rat anti-idiotypic MoAb (9G4[44]). Although we successfully enriched for VH4-34 encoded antibodies, the Fab/phage were not specific for Rh(D) and displayed serological characteristics similar to those of cold agglutinins (data not shown). We are currently examining a μ phage display library from the same donor to compare gene usage.

Rh(D) Epitopes and Significance of Antibody Sequences

Since the initial report by Argall et al[45] in 1953, it has been recognized that rare individuals who type as Rh(D)-positive can produce allo-anti-Rh(D) antibodies in response to Rh(D) immunization by transfusion or pregnancy. This phenomenon was explained by hypothesizing that the Rh(D) antigen is a mosaic structure and that these individuals were producing alloantibodies to parts of the mosaic they lack. By systematically examining patterns of reactivity between their cells and sera, RBCs expressing partial Rh(D) antigens were divided into categories, each presumed to have a different abnormality in their Rh(D) antigen. Through the subsequent use of index panels of monoclonal anti-Rh(D) antibodies, a series of epitopes were defined of which the number and combination varied from one Rh(D) category to another. As new monoclonals were produced, their reactivity profiles against these partial Rh(D) RBCs became the standard method for determining Rh(D) antibody epitope specificity. Molecular analyses of partial Rh(D) phenotypes have shown that the Rh(D) genes in these individuals have either undergone intergenic recombination with the highly homologous Rh(CE) gene or, less commonly, have sustained point mutation(s).[46]

As noted earlier, to investigate the topological relationships among Rh(D) epitopes, Gorick et al[37] performed competition experiments with Rh(D) MoAbs and observed varying degrees of inhibition. These results, when combined with those of Lomas et al,[6] suggested a model for Rh(D) in which epitopes are spatially distinct yet demonstrate a certain degree of overlap, as shown in Fig 10A. This model explained how antibodies to two different Rh(D) epitopes (in this case epD2 and epD3) could inhibit each other's binding to wild-type Rh(D) and how a change in the structure of Rh(D) in category VI RBCs (asterisk, Fig 10A) would cause the loss of epD2. However, based on this concept of Rh(D) epitopes as distinct domains, we would expect that antibodies against different epitopes of Rh(D) would be structurally and genetically distinct as well. Thus, it was surprising that our anti-Rh(D) clones demonstrated such marked restriction in gene usage. For example, only two superspecies of V_H genes were used despite specificities for 4 of the original 6 Rh(D) epitopes described by Lomas et al.[6] Furthermore, multiple specificities could arise from a single heavy chain depending on the light chain with which it was paired (eg, E1 with M2, M3, L3, or L4). In addition, other clones repeatedly

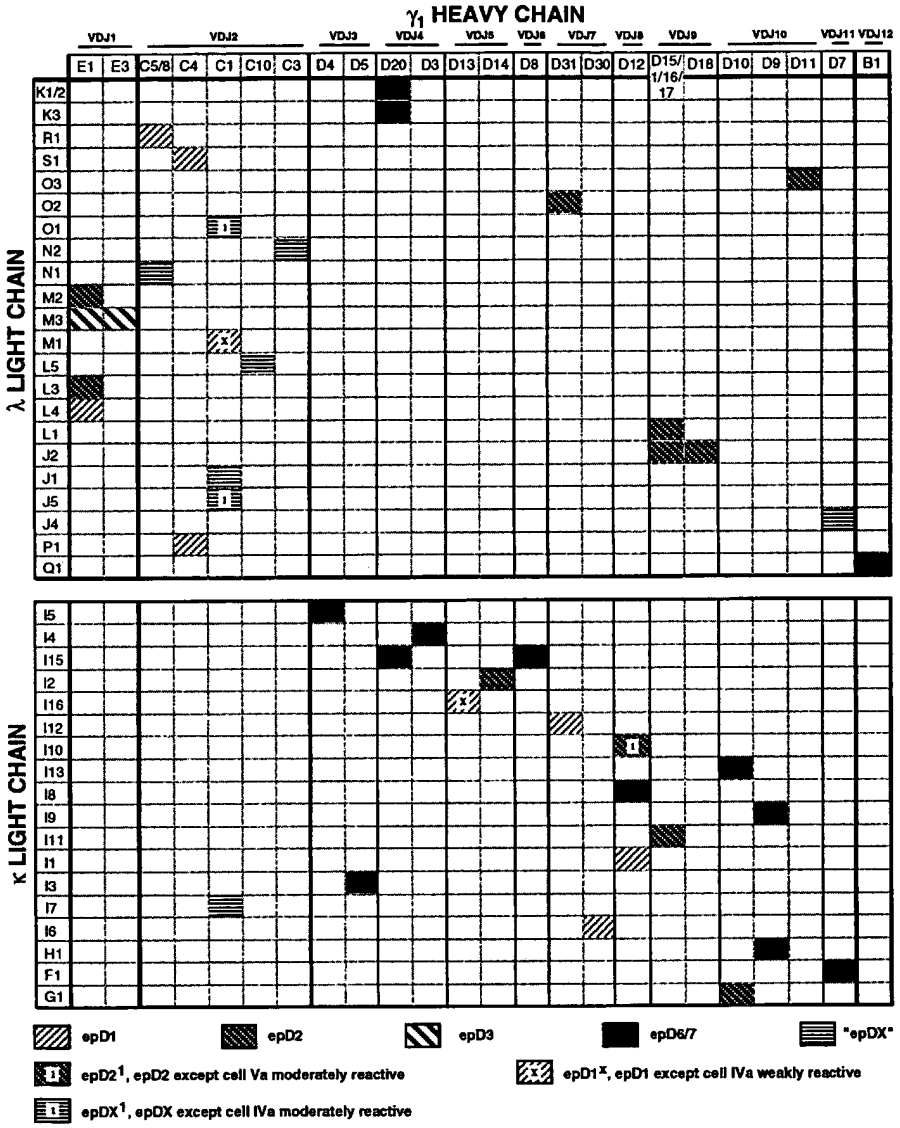


Fig 8. Matrix illustrating the genetic composition and epitope specificity of anti-Rh(D) antibodies. The horizontal axis represents the unique γ_1 heavy chains and the vertical axis represents the unique λ and κ light chains (based on amino acid sequence). A shaded pattern at the intersection of a heavy chain/light chain pair indicates the Rh(D) epitope specificity observed for that Fab/phage antibody. A few clones gave mixed patterns of reactivity, as shown (see text). The order of heavy chains (left to right) and light chains (top to bottom) was determined by the multiple alignment of amino acid sequences as in Figs 2, 4, and 5. Note that heavy chains D1, D15, D16, and D17, although differing in nucleotide sequence, have the identical amino acid sequences and thus comprise a single column. Similarly, heavy chains C5 and C8 and λ light chains K1 and K2 encode the same proteins. The pairings of these 28 heavy and 41 light chain nucleotide gene segments, which produced 53 unique Fab transcripts, encoded 43 different Fab proteins, as indicated in the matrix.

demonstrated variable weak reactivity against certain Rh(D) category RBCs that would affect the epitope specificities to which they were assigned (eg, C1 with O1, M1, or J5).

Several hypotheses could account for these findings. The most simplistic interpretation is that the heavy chain does not directly interact with the antigen, but rather is responsible for bringing the antibody in close proximity with the antigen. The specific interactions between the light chain and the antigen would then determine the epitope specificity for that antibody. In this regard, our data are consistent with the observations of Boucher et al[10] on the relative cationic nature of anti-Rh(D) heavy chains. However, because we found that light chains become cationic during somatic mutation, the charge of the entire antibody may play a role in its ability to bind, resulting in the selection and expansion of particular B-cell clones.

A more compelling hypothesis is that Rh(D) epitopes do not differ spatially but differ only in the number and arrangement of contact residues presented (Fig 10B). In other words, the footprints of most, if not all, anti-Rh(D) antibodies are essentially identical to one another. The genetic events that produce partial Rh(D) molecules result in the loss of certain critical key points of contact necessary for some antibodies to bind; alternatively, they result in the formation of new structures that interfere with the binding of other anti-Rh(D) Igs. For example, the introduction of a ledge in Rh(D) category VI cells (asterisk, Fig 10B) does not interfere with the binding of an anti-epD3 antibody, but does prevent the binding of anti-epD2. Therefore, category VI RBCs are said to have epD3 but lack epD2.

This model is consistent with our inhibition experiments (Fig 9) and with those of Gorick et al[37] and offers an explanation for the marked restriction in heavy chain gene usage. It also reconciles a mechanism by which one heavy chain (eg, E1) can confer binding to multiple epitopes and why some of our recombinant anti-Rh(D) antibodies, as well as some convention-

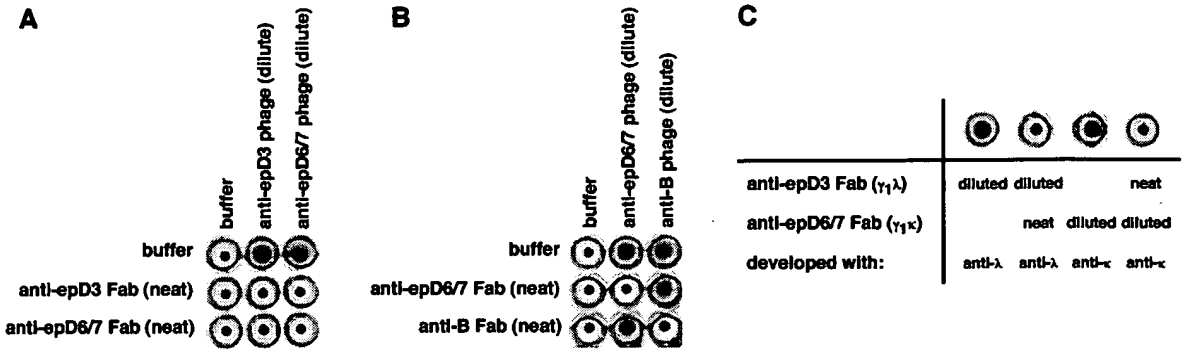


Fig 9. Inhibition studies with recombinant anti-Rh(D) antibodies. Panels show results of representative experiments demonstrating the mutual inhibition of antibodies directed at 2 different Rh(D) epitopes (in this example, epD3 and epD6/7; A and C), but not between an Rh(D) antibody and an unrelated recombinant anti-RBC antibody (an anti-blood group B antibody; B). In (A), Rh(D)-positive RBCs were incubated with soluble Fabs only, phage-displayed Fabs only, or combinations of the two, as indicated. In (B), Rh(D)-positive RBCs that were blood group B were used. After washing, RBCs were resuspended in anti-M13 antibody and assessed for agglutination induced by phage-displayed Fabs. Soluble Fabs were used full-strength, whereas Fab/phage preparations were present in limiting amounts to increase the sensitivity of the inhibition assay (see the Materials and Methods). In (C), mutual inhibition of epD3 and epD6/7 anti-Rh(D) antibodies was demonstrated with Rh(D)-positive RBCs, $\gamma_1\kappa$ and $\gamma_1\lambda$ soluble Fabs, and light chain isotype-specific antisera (see text for details). In these examples, the anti-epD3 and anti-epD6/7 antibodies were clones E1/M3 and D5/I3, respectively. The anti-blood group B antibody was isolated from an IgG phage display library made from the splenic B cells of a blood group O donor.[61]

ally produced monoclonals,[36] display variable reactivity against certain categories of partial Rh(D) RBCs. From the antigen's perspective, this model explains how a single point mutation in Rh(D) can result in the loss of multiple Rh(D) epitopes (such as T283I in category HMi RBCs[47]) and how the residues associated with the expression of some epitopes appear to be distributed among nearly all the extracellular loops of Rh(D).[48] It also provides an understanding as to how ≥37 epitopes can fit on the relatively small extracellularly exposed surface of the Rh(D) molecule.[3]

This concept of coincident epitopes is best exemplified by comparing the E1/M2 and E1/M3 clones. The only difference between the reactivity of E1/M2 and E1/M3 is the ability of the latter antibody to agglutinate Rh(D) category VI cells (Fig 7). Hence, E1/M2 is classified as an anti-epD2 and E1/M3 as an anti-epD3 antibody. Light chains M2 and M3 differ by only 3 residues: D82A, G95aA, and W96V (Fig 5). Therefore, some combination of these residues confers reactivity against category VI cells. In other words, epD2 and epD3, as seen by the E1/M2 and E1/M3 antibodies, differ by the binding constraints

Table 1. Comparison of Current IgG Fab/Phage Library-Derived Anti-Rh(D) MoAbs With Those Previously Produced by Conventional Tissue Culture Methods

Attribute	Previously Published*	Current Study	
Heavy Chains		(by clone)†	(by VDJ)
VH3 family derived	12/21 (57%)	28/28 (100%)	12/12 (100%)
VH3-33 superspecies‡/VH3	10/12 (83%)	26/28 (93%)	11/12 (92%)
VH3-33/VH3	9/12 (75%)	19/28 (68%)	9/12 (75%)
VH3-21/VH3	1/12 (8%)	2/28 (7%)	1/12 (8%)
VH4-34 derived	2/21 (10%)	0/28 (0%)	0/12 (0%)
JH6 usage	15/21 (71%)	9/28 (32%)	5/12 (42%)
CDR3 length	16-20 (28§)	15-19	
κ Light Chains			
Vκ1 family derived/total κ	8/12 (67%)	17/18 (94%)	
Jκ1 usage/total κ	4/12 (33%)	6/18 (33%)	
Jκ2 usage/total κ	4/12 (33%)	6/18 (33%)	
λ Light Chains			
Vλ1 family derived/total λ	2/8 (25%)	12/23 (52%)	
Vλ3 family derived/total λ	5/8 (63%)	5/23 (22%)	
DPL16 derived/Vλ3 family	3/5 (60%)	4/5 (80%)	
Jλ2Vasicek usage/total λ	6/8 (75%)	23/23 (100%)	

*Compiled from a total of 21 sequences of IgG anti-Rh(D) antibodies isolated from multiple subjects originally published by Bye et al,[25] Hughes-Jones et al,[23] Chouchane et al,[24] and Boucher et al[10] and available from Genbank. One light chain (Oak-3)[25] was not available in Genbank and was not included in the assessment.

†For heavy chains, the left column tabulates each clone separately and the right column tabulates clones on the basis of shared V-D-J joining regions.

‡VH3-33 superspecies defined as the group of VH3 family germline genes comprising VH3-33, VH3-30, and VH30.3.

§CDR3 length outlier.

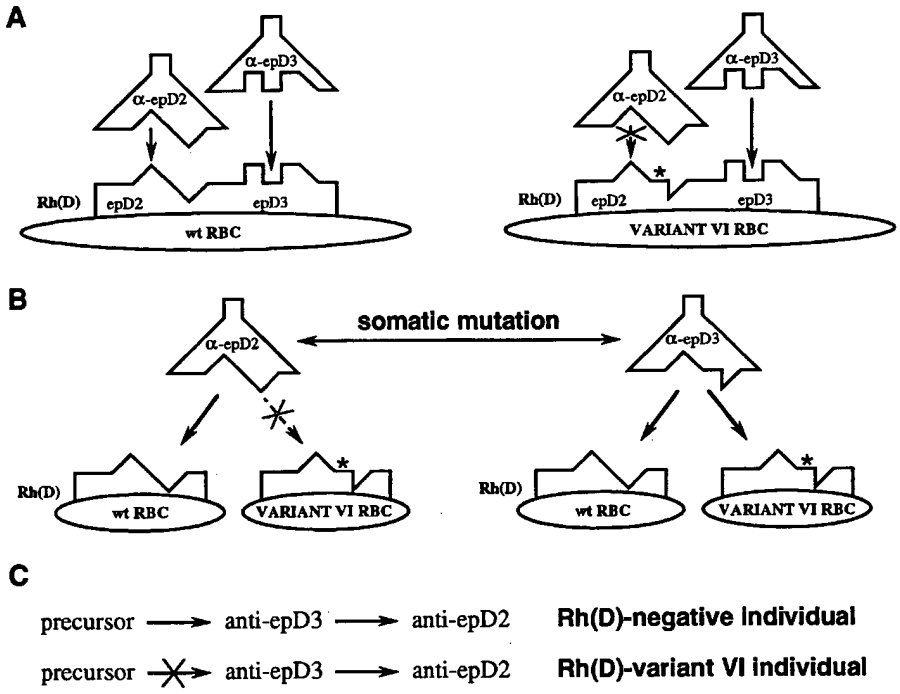


Fig 10. Conventional (A) and proposed (B) models for Rh(D) antigen/antibody binding. Note that the predicted combining sites and genetic relationships between antibodies differ between the two models. (C) If antibodies directed at different Rh(D) epitopes are clonally related, then the expressed repertoire will differ between Rh(D)-negative and partial Rh(D) individuals (see text for discussion).

imposed by at most three mutations. If the model depicted in Fig 10A were correct and the epitopes were independent, these mutations would have to cause enough structural alteration in the antibody combining site so that a completely separate epitope on the same antigen would be recognized. It would seem unlikely that these 3 mutations could cause such a change, especially given the lack of internal homology domains in Rh(D). Thus, we conclude that it is far more plausible that the footprints of these 2 antibodies are essentially identical and that one or more of these mutations (eg, the tryptophan in CDR3 of M2) prevent(s) the interaction of E1/M2 with category VI RBCs. Because other clones demonstrate that the light chain can confer specificity against epD1, epD2, or epD3 (with the E1 heavy chain); epD1 or epDX (with C5); and epD1, epD2, and epD6/7 (with D12), we suggest that all 5 of these epitopes have similar antibody combining sites.

Immunologic and Clinical Implications of Proposed Model

The model depicted in Fig 10B leads to additional predictions concerning the Rh(D) immune response beyond simply clarifying what is meant by an Rh(D) epitope. It is commonly stated in the transfusion medicine literature that individuals whose RBCs express partial Rh(D) antigens are free to make antibodies to the Rh(D) epitopes they lack.[34] Therefore, an individual who produces category VI RBCs should be able to make anti-epD2 but not anti-epD3. If these epitopes were truly independent, then the immune repertoire of the anti-epD2 antibodies made by a category VI individual would be similar to those produced by an Rh(D)-negative person. However, to the immune system, epD2 and epD3 are not independent. We postulate that the somatic mutation of an anti-epD3 antibody can change its fine specificity to that of epD2 (or vice versa, Fig 10C). Suppose that the preferred way of making an anti-epD2 antibody is to go through an anti-epD3 intermediate. To an Rh(D)-negative individual, this process can take place unimpeded. However, in a category VI individual, this route would be unfavorable because an anti-epD3 antibody would be self-reactive. As a result, such an individual would have to make anti-epD2 antibodies by following alternative routes or by tolerating some degree of autoreactivity in the process. With respect to the latter point, it is of interest to note that a transient production of auto-anti-Rh(D) frequently precedes or accompanies the early production of allo-anti-Rh(D) in individuals who express partial Rh(D) antigens.[49-54] We would predict, therefore, that the anti-epD2 antibodies from a category VI individual would be different in composition (ie, gene usage) and quite possibly quantitatively depressed as compared with an Rh(D)-negative individual. This may be analogous to the antibodies of the ABO blood group system in which it has been observed that anti-A and anti-B titers in blood group O individuals are significantly higher than in blood group B or A individuals, respectively.[55] Blood group O individuals are unconstrained in creating their anti-A and anti-B immune repertoires, whereas individuals who produce A or B antigens (2 nearly identical structures) must do so in a manner that avoids self-reactivity.

In the case of antibodies E1/M2 and E1/M3, they appear to have arisen from a common precursor B cell rather than directly from each other (Fig 5). To test the framework of our hypothesis, ie, somatic mutation resulting in epitope migration of an antibody, we are constructing the precursors and potential intermediates between the M2 and M3 light chains and will then determine what Rh(D) epitope specificities (if any) they express. This concept of epitope migration has been previously reported for murine anti-cryptococcal[56] and anti-type II collagen[57] antibodies.

If the proposed model for Rh(D) epitopes is correct, then the

question of the number of epitopes may be obsolete. There may be as many epitopes as can be differentiated by the number of cell categories, ie, 2^n epitopes, where n is the number of distinct partial Rh(D) RBCs. A more important question is the interrelationships between the various epitopes. For example, are some epitopes further away than others—not in the topological sense, but in terms of the number of mutational hits an antibody needs to receive to change its serologic reactivity. Furthermore, does the humoral immune response in a partial Rh(D) individual differ from that in an Rh(D)-negative individual in the manner predicted by this model? One may find that allo-anti-Rh(D) antibodies made by partial Rh(D) individuals are not as clinically significant, ie, capable of inducing hemolysis. This may explain why hemolytic disease of the newborn due to anti-Rh(D) produced by pregnant individuals with partial Rh(D) phenotypes is so rare even when taking into account the low prevalence of the partial Rh(D) phenotypes.[34] A better understanding of the immune response to Rh(D) in these patients may alleviate concerns regarding the need to identify such individuals to ensure that they only receive Rh(D)-negative blood products for transfusion and Rh(D)-immune globulin during pregnancy.[58] Furthermore, with respect to the design of recombinant Rh(D)-immune globulin for use in Rh(D)-negative patients, it may not be necessary to formulate cocktails of MoAbs containing multiple Rh(D) epitope specificities.

In summary, we have studied the genetic and immunological properties of a large array of anti-Rh(D) antibodies to elucidate this clinically significant human immune response on a molecular level. Our results show that anti-Rh(D) antibodies display a high degree of structural relatedness and the ability to inhibit each other's binding despite differences in epitope specificity. These findings suggest that Rh(D) epitopes are not spatially distinct and that Rh(D) antibodies may undergo epitope migration as a result of somatic mutation. The end result is that the prevalence of certain anti-Rh(D) specificities in the immune repertoire may be a function not only of what epitopes an individual lacks, but of the number of accessible pathways that the individual's immune system can use that avoid self-reactivity. This process may be a general feature of human immune responses to other clinically significant, closely related epitopes.

ACKNOWLEDGMENT

The authors thank Shari Russell and Vatinee Bunya for their excellent technical assistance and Christine Lomas-Francis, Marion Reid, Marilyn Moulds, and Peggy Spruell for providing samples of Rh(D) variant RBCs.

APPENDIX

Genbank accession numbers for anti-Rh(D) heavy chains are as follows: B01, AF044419; C01, AF044420; C03, AF044421; C04, AF044422; C05, AF044423; C08, AF044424; C10, AF044425; D01, AF044426; D03; AF044427; D04, AF044428; D05, AF044429; D07, AF044430; D08, AF044431; D09, AF044432; D10, AF044433; D11, AF044434; D12, AF044435; D13, AF044436; D14, AF044437; D15, AF044438; D16, AF044439; D17, AF044440; D18, AF044441; D20, AF044442; D30, AF044443; D31, AF044444; E01, AF044445; E03, AF044446. Genbank accession numbers for antiRh(D) κ light chains are as follows: F01, AF044447; G01, AF044448; H01, AF044449; I01, AF044450; I02, AF044451; I03, AF044452; I04, AF044453; I05, AF044454; I06, AF044455; I07, AF044456; I08, AF044457; I09, AF044458; I10, AF044459; I11, AF044460; I12, AF044461; I13, AF044462; I15, AF044463; I16, AF044464. Genbank accession numbers for anti-Rh(D) λ light chains are as follows: J01, AF044465; J02, AF044466; J04, AF044467; J06, AF044468; K01, AF044469; K02, AF044470; K03, AF044471; L01, AF044472; L03, AF044473; L04, AF044474; L05, AF044475; M01, AF044476; M02, AF044477; M03, AF044478; N01, AF044479; N02, AF044480; O01, AF044481; O02, AF044482; O03, AF044483; P01, AF044484; Q01, AF044485; R01, AF044486; S01, AF044487.

REFERENCES

1. Bowman JM: Hemolytic disease of the newborn, in Garratty G (ed): Immunobiology of Transfusion Medicine. New York, NY, Dekker, 1994, p 553

2. Shirey R, Ness P: Alloimmunization to blood group antigens, in Anderson KC, Ness PM (eds): Scientific Basis of Transfusion Medicine: Implications for Clinical Practice. Philadelphia, PA, Saunders, 1994, p 507

3. Scott M: Rh serology—Coordinator's report. Transfus Clin Biol 3:333, 1996

4. Moore S, Woodrow CF, McClelland DB: Isolation of membrane components associated with human red cell antigens Rh(D), (c), (E) and Fy^a. Nature 295:529, 1982

5. Gahmberg CG: Molecular characterization of the human red cell Rh_o(D) antigen. EMBO J 2:223, 1983

6. Lomas C, Tippett P, Thompson KM, Melamed MD, Hughes-Jones NC: Demonstration of seven epitopes on the Rh antigen D using human monoclonal anti-D antibodies and red cells from D categories. Vox Sang 57:261, 1989

7. Bloy C, Hermand P, Blanchard D, Cherif-Zahar B, Goossens D, Cartron J-P: Surface orientation and antigenic properties of Rh and LW polypeptides of the human erythrocyte membrane. J Biol Chem 265:21482, 1990

8. Le van Kim C, Mouro I, Cherif-Zahar B, Raynal V, Cherrier C, Cartron JP, Colin Y: Molecular cloning and primary structure of the human blood group RhD polypeptide. Proc Natl Acad Sci USA 89:10925, 1992

9. Arce MA, Thompson ES, Wagner S, Coyne KE, Ferdman BA, Lublin DM: Molecular cloning of RhD cDNA derived from a gene present in RhD-positive, but not RhD-negative individuals. Blood 82:651, 1993

10. Boucher G, Broly H, Lemieux R: Restricted use of cationic germline V_H gene segments in human Rh(D) red cell antibodies. Blood 89:3277, 1997

11. Stein LD, Sigal NH: Limiting dilution analysis of Epstein Barr virus-induced immunoglobulin production. Cell Immunol 79:309, 1983

12. Brown NA, Miller G: Immunoglobulin expression by human B lymphocytes clonally transformed by EBV. J Immunol 128:24, 1982

13. Roder JC, Cole SPC, Kozbor D: The EBV-hybridoma technique. Methods Enzymol 121:140, 1986

14. Goossens D, Champomier F, Rouger P, Salmon C: Human monoclonal antibodies against blood group antigens. J Immunol Methods 101:193, 1987

15. Casali P, Burastero SE, Nakamura M, Inghirami G, Notkins AL: Human lymphocytes making rheumatoid factor and antibodies to ssDNA belong to Leu-1 positive B-cell subset. Science 236:77, 1987

16. Croce CM, Shander M, Martinis J: Chromosomal location of the gene for human immunoglobulin heavy chains. Proc Natl Acad Sci USA 76:3416, 1979

17. Croce CM, Shander M, Martinis J: Preferential retention of human chromosome 14 in mouse x human B cell hybrids. Eur J Immunol 10:486, 1980

18. Burton DR, Barbas CF: Human antibodies from combinatorial libraries. Adv Immunol 57:191, 1994

19. Winter G, Griffiths AD, Hawkins RE, Hoogenboom HR: Making

antibodies by phage display technology. Annu Rev Immunol 12:433, 1994

20. Siegel DL: New approaches to monoclonal antibody production, in Garratty G (ed): Applications of Molecular Biology to Blood Transfusion Medicine. Bethesda, MD, AABB, 1997, p 73

21. Siegel DL, Silberstein LE: Expression and characterization of recombinant anti-Rh(D) antibodies on filamentous phage: A model system for isolating human red blood cell antibodies by repertoire cloning. Blood 83:2334, 1994

22. Siegel DL, Chang TY, Russell SL, Bunya VY: Isolation of cell surface-specific human monoclonal antibodies using phage display and magnetically-activated cell sorting: Applications in immunohematology. J Immunol Methods 206:73, 1997

23. Hughes-Jones NC, Bye JM, Beale D, Coadwell J: Nucleotide sequences and three-dimensional modeling of the V_H and V_L domains of two human monoclonal antibodies specific for the D antigen of the human Rh-blood-group-system. Biochem J 268:135, 1990

24. Chouchane L, Van Spronsen A, Breyer J, Guglielmi P, Strosberg AD: Molecular characterization of a human anti-Rh(D) antibody with a DH segment encoded by a germ-line sequence. Eur J Biochem 207:1115, 1992

25. Bye JM, Carter C, Cui Y, Gorick BD, Songsivilai S, Winter G, Hughes-Jones NC, Marks JD: Germline variable region gene segment derivation of human monoclonal anti-Rh(D) antibodies. J Clin Invest 90:2481, 1992

26. Barbas CF, Lerner RA: Combinatorial immunoglobulin libraries on the surface of phage (Phabs): Rapid selection of antigen-specific Fabs. Methods: A Companion to Methods Enzymol 2:119, 1991

27. Roben PW, Salem AN, Silverman GJ: VH3 family antibodies bind domain D of staphylococcal protein A. J Immunol 154:6437, 1995

28. Althaus H-H, Müller W: DNAPLOT. http://www.mrc_cpe.cam-.ac.uk/imt_doc/DNAsearch.html (internet), 1996

29. Tomlinson IM, Williams SC, Corbett SJ, Cox JPL, Winter G: V Base Directory of Human V Gene Sequences. http://www.mrc_cpe.cam.ac.uk/imt_doc/vbase_home_page.html (internet), 1996

30. Siegel D, Russell S, Chang T: A plethora of human monoclonal anti-Rh antibodies isolated using phage display and magnetically-activated cell sorting. Transfusion 36:37S, 1996

31. Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G: Structural repertoire of the human VH segments. J Mol Biol 227:799, 1992

32. Dörner T, Brezinschek H-P, Brezinschek R, Foster S, Domiati-Saad R, Lipsky P: Analysis of the frequency and pattern of somatic mutations within nonproductively rearranged human variable heavy chain genes. J Immunol 158:2779, 1997

33. Stollar BD: The expressed heavy chain V gene repertoire of circulating B cells in normal adults. Ann NY Acad Sci 764:547, 1995

34. Mollison PL, Engelfreit CP, Contreras M: Blood Transfusion in Clinical Medicine (ed 9). Oxford, UK, Blackwell Scientific, 1993

35. Tippett P, Moore A: Monoclonal antibodies against Rh and Rh related antigens. J Immunogenet 17:309, 1990

36. Tippett P, Lomas-Francis C, Wallace M: The Rh antigen D: Partial D antigens and associated low incidence antigens. Vox Sang 70:123, 1996

37. Gorick BD, Thompson KM, Melamad MD, Hughes-Jones NC: Three epitopes on the human Rh antigen D recognized by ^{125}I-labeled human monoclonal IgG antibodies. Vox Sang 55:165, 1988

38. Abelson N, Rawson A: Studies of blood group antibodies. II. Fractionation of Rh antibodies by anion-cation cellulose exchange chromatography. J Immunol 83:49, 1959

39. Frame M, Mollison P: Charge differences between human IgG isoantibodies associated with specificity. Immunology 16:277, 1969

40. Silberstein LE, Jefferies LC, Goldman J, Friedman DF, Moore JS, Nowell PC, Roelcke D, Pruzanski W, Roudier J, Silverman GJ: Variable region gene analysis of pathologic human autoantibodies to the related i and I red blood cell antigens. Blood 78:2372, 1991

41. Pascuel V, Victor K, Leisz D, Spellerberg MB, Hamblin TJ, Thompson KM, Rannden I, Natvig J, Capra JD, Stevenson FK: Nucleotide sequence analysis of the V regions of two IgM cold agglutinins: Evidence that the VH4-21 gene segment is responsible for the major cross-reactive idiotype. J Immunol 146:4385, 1991

42. Silverman G, Goni F, Fernandez J: Distinct patterns of heavy chain variable region subgroup use by human monoclonal autoantibodies of different specificity. J Exp Med 168:2361, 1988

43. Thompson KM, Sutherland J, Barden G, Melamed MD, Randen I, Natvig JB, Pascual V, Capra JD, Stevenson FK: Human monoclonal antibodies against blood group antigens preferentially express a V_H4-21 variable region gene-associated epitope. Scand J Immunol 34:509, 1991

44. Stevenson FK, Smith GJ, North J, Hamblin TJ, Glennie MJ: Identification of normal B-cell counterparts of neoplastic cells which secrete cold agglutinins of anti-I and anti-i specificity. Br J Haematol 72:9, 1989

45. Argall C, Ball J, Trentleman E: Presence of anti-D antibody in the serum of a D^u patient. J Lab Clin Med 41:895, 1953

46. Cartron J, Rouillac C, Le Van Kim C, Mouro I, Colin Y: Tentative model for the mapping of D epitopes on the RhD polypeptide. Transfus Clin Biol 3:497, 1996

47. Liu W, Jones J, Scott M, Voak D, Avent N: Molecular analysis of two D-variants, DHMi and DHMii. Trans Med 6:21, 1996 (suppl 2, abstr)

48. Scott M, Voak D, Jones J, Avent N, Liu W, Hughes-Jones N, Sonneborn H: A structural model for 30 Rh D epitopes based on serological and DNA sequence data from partial D phenotypes. Transfus Clin Biol 3:391, 1996

49. Chown B, Kaita H, Lewis M, Roy B, Wyatt L: A "D-positive" man who produced anti-D. Vox Sang 8:420, 1963

50. Macpherson C, Stevenson T, Gayton J: Anti-D in a D-positive man, with positive direct Coombs test and normal red cell survival. Am J Clin Pathol 45:748, 1966

51. Beard M, Pemberton J, Blagdon J, Jenkins W: Rh immunization following incompatible blood transfusion and a possible long-term complication of anti-D immunoglobulin therapy. Med Genet 8:317, 1971

52. Cook I: Primary rhesus immunization of male volunteers. Br J Haematol 20:369, 1971

53. Holland P, Gralnick M: A new category? Auto-anti-D with complete D antigen [sic] but no overt hemolysis. Transfusion 13:363, 1973 (abstr)

54. Issitt PD: Applied Blood Group Serology (ed 3). Miami, FL, Montgomery Scientific, 1985

55. Ichikawa Y: A study of the iso-agglutinin titers in the sera of Austrailian subjects (white). Jpn J Med Sci Biol 12:1, 1959

56. Mukherjee J, Nussbaum G, Scharff MD, Casadevall A: Protective and nonprotective monoclonal antibodies to *Cryptococcus neoformans* originating from one B cell. J Exp Med 181:405, 1995

57. Mo JA, Holmdahl R: The B cell response to autologous type II collagen. J Immunol 157:2440, 1996

58. Jones J, Scott ML, Voak D: Monoclonal anti-D specificity and Rh D structure: Criteria for selection of monoclonal anti-D reagents for routine typing of patients and donors. Trans Med 5:171, 1995

59. Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C: Sequences of Proteins of Immunological Interest (ed 5). US Department of Health and Human Services, NIH Publication number 91-3242, Bethesda, MD, 1991

60. Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G: The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. J Mol Biol 227:776, 1992

61. Chang T, Siegel D: Isolation of a highly mutated human anti-B antibody from a phage-display library. Transfusion 37:89S, 1997 (abstr)